

- Firstly, the authors would like to thank the reviewers for their comments, almost all of which have resulted in changes to the revised manuscript. We feel that this has very much improved the work and should make it more amenable to use by the wider research community.
- A new, edited and updated manuscript has been prepared within which all new material has been highlighted by the inclusion of blue text. Other notable alterations are as follows
  1. Several new references have been included
  2. Sections 5.2.2 (Vertical nitrate profiles) and 5.2.3 (Ocean productivity) are new
  3. The far right column in Table 3 is new
  4. Table 4 is new
  5. Figures 4, 8, 9 and 15-18 are new
  6. Figure 5 has been altered to highlight the top 3 scoring land-surface ensemble members.
- We now move on to addressing the two anonymous referees' comments in turn.

#### **Referee comment #1**

- In light of the referee's comment concerning the marine carbon cycle; depth-latitude plots of nitrate concentration have been added for the Pacific-basin, the Atlantic-basin and globally (Figures 15-17). In addition to this, another new figure (Figure 18) has been added which shows the ocean productivity data for the generation 3 and 4 models.
- Currently the biogeochemical capability of FAMOUS does not include oxygen in the ocean biogeochemical scheme and therefore no results can be shown. The authors however have added a paragraph in the "Discussions, conclusions and future work" section detailing their future plans with this modelling framework, which includes the addition of oxygen as a prognostic variable in the ocean.
- In the configuration of FAMOUS presented in this work, the atmospheric CO<sub>2</sub> concentration is held fixed and therefore it was felt that the inclusion of surface CO<sub>2</sub> fluxes did not yield information which would be of use to the general model developer readership (the intended audience). Work is currently underway to make the atmospheric carbon dioxide concentration variable and this will be addressed in a forthcoming publication, as indicated in the "Discussions, conclusions and future work" section.

#### **Referee comment #2**

1. Points (a) and (b) will be addressed separately:
  - a)
    - i. More information has been included regarding the structure and workings of the dynamic vegetation model, TRIFFID, and its relationship to the MOSES

- surface exchange scheme. The interested reader is additionally referred to the original references for these models for further information.
- ii. It is clear that the use of the word 'equilibrium' to describe the temporal evolution of the dynamic vegetation in the ensemble members has caused confusion and this has been replaced by the word 'accelerated'. The relevant paragraph on page 3093 of the original draft has also been reworded/clarified to remove certain terms which will be unfamiliar to the general readership.
- b) The description of the HadOCC (ocean carbon cycle) model has been enlarged (as suggested) to give the general reader more information and the more interested reader is referred to the specific section of the original documentation paper (Palmer and Totterdell, 2001) where a detailed and complete model description is given.
2. As the referee correctly states, global vegetation will certainly take longer than 200 years to come into equilibrium and this is the reason for the use of the accelerated, asynchronously coupled, scheme in this work. As stated above (point 1-a-ii) the authors have reworded the section concerning the 'accelerated' dynamic vegetation scheme.
  3. The question pertains to why 20 parameters were varied in the ocean ensemble whilst only 7 were varied in the land surface ensemble. The parameters in the ocean ensemble were grouped together and co-varied according to their compartmentalisation in the model (phytoplankton parameters, zooplankton parameters, etc). As a result of this, there are actually only 5 freely varying parameters in the ocean ensemble, which is indicated in the far right column of Table 2. The authors have made some changes to the wording of section 3 of the manuscript and hope that this has now been made clear.
  4. Points (a) and (b) will be addressed separately:
    - a)
      - i. An extra figure has been added (Figure 4) which shows the PFT distributions of the observations, of HadCM3 (FAMOUS' parent model), the 3 top ' $\alpha 7$ ' simulations, and 3 further ensemble members sampled to illustrate steadily decreasing agreement over the Amazon region, across the 100 member ensemble. It is shown that although there are notable biases present in the top 3  $\alpha 7$  simulations, the results for HadCM3 still contain some significant biases. The top 3  $\alpha 7$  simulations are also shown to be very similar to each other globally, indicating that the resulting simulated vegetation distribution is robust.
      - ii. The 3 best performing  $\alpha 7$  simulations indicated in Figure 4(c-e) are also highlighted with different shaped symbols in an updated Figure 5 (previously figure 3) as suggested by the referee: *"If the results of the best 7 [we show the best 3] are shown pictorially, these should then be identified individually in Figure 3. This would be most informative and interesting for the reader."*
      - iii. As regards the utility of the observational dataset, the referee states that *"Equatorial Africa, for example, appears to be dominated by C3 and C4 vegetation whereas surely some sort of rainforest should dominate?"*. The observations (Figure 1) do show broadleaf forest coverage in Equatorial Africa in regions (purple) covered by countries such as Cameroon, Central African Republic and The Democratic Republic of the Congo. The original dataset (Loveland et al., 2000) contains 17 surface types and since the dataset used in Figure 1 was 'downscaled' to just 5 will inevitably mean that

there is some loss of information. The original 1km dataset was also regridded to the resolution of the FAMOUS atmosphere (i.e. approximately 280km x 420km), which will lead to important forest cover on smaller scale not being evident in this dataset, in some regions, particularly along the West African southern coast.

- iv. The reasons for focussing on the Amazon region have been significantly expounded in the text. In brief, the Amazon was chosen, partly because it is known to have a significant influence on global climate and carbon budget. Secondly, it is the only large area of the globe which, in the observations, has a fractional coverage of the dominant PFT close to unity (with the exception of the Sahara, which is clearly devoid of vegetation), which is an emergent property of the simulations (Figure 2).
- v. In response to *"The authors should also consider offering some discussion about how well the key parameters can be constrained by data in context of the ranges shown in Figure 3, thereby addressing the issue of uncertainty."* we have expanded our discussion of the role model-observational comparisons can play in constraining GCM parameters (Section 6).

b)

- i. The authors have added a further figure (Figure 9) which shows the performance of five different ensemble members with decreasing Arcsine Mielke scores (from a maximum of 0.71 to a minimum of 0.04). The addition of this figure is in response to the following comment *"The authors should consider showing, for example, the predicted nitrate distributions for several of the other ensemble members, e.g. the best one and a random selection of the others, to give the reader a feel for the degree of variability here."*
- ii. It has been newly demonstrated (far-right column in Table 3 and Figure 8) that three of the five perturbed parameters are not appreciably constrained by the top six scoring ensemble members (essentially indistinguishable by eye in Figure 7), that is, the parameters in these six simulations cover almost the whole range of parameter space. However, the parameter controlling the C:N ratios is confined to approximately the top two-thirds of its explored parameter range and the phytoplankton parameters are constrained to roughly the top 25%. This is precisely analogous to the result for the land surface ensemble which showed that the parameter  $f_0$  is constrained to high values within its perturbed range by fitting to the dominant PFT found in the Amazon region. Further consideration of the top ten performing ensemble members gives the same result as when only the top six are considered which adds extra weight to the assertion that phytoplankton-specific parameters, in particular, are well constrained by this work.
- iii. In addition to the surface nitrate distributions already shown, the authors have added validation plots for (zonal mean) nitrate as a function of depth globally and for the Pacific and Atlantic basins (Figures 15-17). In addition to this, simulated ocean productivity data has been compared to observations derived from satellite-observed chlorophyll concentrations<sup>1</sup> (Figure 18).
- iv. As suggested by the referee, a section has been added discussing the physiological validity of the identified parameter ranges and this is made in particular regard to the study of Scott et al. (2011) who have performed a similar, albeit reduced complexity, perturbed physics ensemble to that considered in this work. It is found that the vast majority of the parameters

---

<sup>1</sup> Photosynthetic Rates Derived from Satellite-Based Chlorophyll Concentration. Michael J. Behrenfeld, Paul G. Falkowski, *Limnology and Oceanography*, Vol. 42, No. 1 (Jan., 1997), pp. 1-20.

perturbed in this work fall within the identified ranges from Scott et al. within which literature-derived ranges are found wherever possible. Both the zooplankton assimilation efficiency ( $\beta_p$ ) and the detrital sinking rate identified from this work fall within the (literature-derived) ranges identified by Scott et al.

- v. The authors have added a few sentences (in the new 'Physiological justification of suitable parameters' section) discussing the validity of each individual parameter's value when several parameters are simultaneously co-varied, as is the case throughout this work.
5. The points made here will be dealt with individually:
- i. Regarding the issue of the state-of-the-art nature of the sub-models used, the HadOCC model is the only biogeochemical model which can currently be coupled to the Hadley Centre GCM ocean. Future work within this modelling framework (indeed, this work has already started) will focus on significantly improving the biogeochemical cycling capabilities of the Hadley Centre model, which is elaborated upon in the text.
  - ii. As regards the use of the combination of the MOSES 2.2 land surface scheme and the TRIFFID dynamic vegetation model; there is a more up-to-date standalone version of the land surface exchange scheme known as JULES<sup>2</sup> (the Joint UK Land Environment Simulator), however, to date MOSES 2.2 is the most sophisticated land surface implemented within a Hadley Centre global climate model configuration. Indeed, the combination of the "2.2" version of MOSES with TRIFFID had not been used before in FAMOUS and this was the reason for the perturbed physics ensemble presented here. The authors acknowledge that this was perhaps not made as clear as it should have been in the original manuscript and additional text has been added to emphasise this.
  - iii. The desired outcome of this work was a fast fully-coupled GCM with dynamic vegetation and an ocean carbon cycle. While FAMOUS does not match some of the process fidelity or high resolution of current GCMs, e.g. Collins et al. (2011), it has been developed explicitly as a coarse resolution model. More information and references have been added (page 1) to expand upon this.
  - iv. As stated on page 3091 of the published manuscript on GMD Discussions, FAMOUS' parent (and higher-resolution) model takes approximately 100 "wall clock" days to run 1000 simulated years, which is considerably too slow for these long simulations to be regularly and usefully run. The authors appreciate that the configuration of the UM used in this work cannot be considered to be "high resolution", however this loss of information inherent to reduced spatial precision (in both ocean and atmosphere/land surface) is a necessary price to pay for a model which can run fast enough to regularly run the deep ocean to dynamic equilibrium. In any case, it is debatable for example (see Figure 4 in the updated manuscript) that the reproduction of the observed dominant PFT distribution in HadCM3 is actually considerably better than that of the FAMOUS, even though the resolution of HadCM3 is twice that of FAMOUS in the horizontal.
  - v. The referee makes the point that perhaps ocean physics is more important than the choice of ocean biogeochemical parameters. In the original

---

<sup>2</sup> Best, M. J., M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards et al. "The Joint UK Land Environment Simulator (JULES), model description—Part 1: energy and water fluxes." *Geoscientific Model Development* 4, no. 3 (2011): 677-699.

manuscript, this point is made with regard to the study of Doney et al. (page 3103) where the authors suggest that a future, more coordinated parameter uncertainty study would be valuable.

- vi. Regarding the suitability of the datasets used, text has been added describing the utility of the Loveland et al. dataset with regard to its advanced quality control techniques and the very high (1km) resolution satellite dataset from which it is derived. As regards the use of the World Ocean Atlas (nitrate) and Behrenfeld & Falkowski (productivity) datasets, the use of these as validation targets were explicitly recommended by the code developers of the HadOCC model at the Met Office Hadley Centre.

# Optimising the FAMOUS climate model: inclusion of global carbon cycling

J. H. T. Williams<sup>1</sup>, R. S. Smith<sup>2</sup>, P. J. Valdes<sup>1</sup>, B. B. Booth<sup>3</sup>, and A. Osprey<sup>2</sup>

<sup>1</sup>BRIDGE, School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK

<sup>2</sup>Department of Meteorology, University of Reading, Reading, RG6 6BB, UK

<sup>3</sup>Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK

*Correspondence to:* J. H. T. Williams (jonny.williams@bristol.ac.uk)

**Abstract.** FAMOUS fills an important role in the hierarchy of climate models, both explicitly resolving atmospheric and oceanic dynamics yet being sufficiently computationally efficient that either very long simulations or large ensembles are possible. An improved set of carbon cycle parameters for this model has been found using a perturbed physics ensemble technique. This is an important step towards building the “Earth System” modelling capability of FAMOUS, which is a reduced resolution, and hence faster running, version of the Hadley Centre Climate model, HadCM3. Two separate 100 member perturbed parameter ensembles were performed; one for the land surface and one for the ocean. The land surface scheme was tested against present day and past representations of vegetation and the ocean ensemble was tested against observations of nitrate. An advantage of using a relatively fast climate model is that a large number of simulations can be run and hence the model parameter space (a large source of climate model uncertainty) can be more thoroughly sampled. This has the associated benefit of being able to assess the sensitivity of model results to changes in each parameter. The climatologies of surface and tropospheric air temperature and precipitation are improved relative to previous versions of FAMOUS. The improved representation of upper atmosphere temperatures is driven by improved ozone concentrations near the tropopause and better upper level winds.

of FAMOUS has a resolution of  $5^\circ \times 7.5^\circ$  (compared to the  $2.5^\circ \times 3.75^\circ$  of HadCM3) and has 11 vertical levels, a significant reduction compared to the 19 in HadCM3. The ocean has twice the resolution of the atmosphere (i.e.  $2.5^\circ \times 3.75^\circ$ ) and 20 vertical levels. HadCM3’s ocean resolution is  $1.25^\circ \times 1.25^\circ$  and also has 20 vertical levels. The atmospheric timestep for FAMOUS is 1 h, twice that of HadCM3, whereas the timestep in the ocean is 12 h, compared to just 1 h for HadCM3. The reduction in model resolution and increase in model timesteps means that FAMOUS runs approximately 10 times faster than its parent model. For example, a 1000 yr, coupled atmosphere-ocean simulation with HadCM3 takes approximately 100 days on 8 processors and generates 1 Tb of model data. An equivalent FAMOUS simulation runs in one tenth of the time and produces one quarter of the amount of output data, due to the lower spatial resolution and longer timesteps in the atmosphere and ocean.

While FAMOUS does not match some of the process fidelity or high resolution of current GCMs, e.g. Collins et al. (2011), it has been developed explicitly as a coarse resolution model. Despite belonging to an older generation of climate models, the atmosphere-ocean component (which is common to FAMOUS and HadCM3) still performs in the best handful of models on larger scale climate measures (Reichler and Kim, 2008; Nishii et al., 2012).

All previously published versions of FAMOUS have used the MOSES (Met Office Surface Exchange Scheme) 1 land surface model (Cox et al., 1999). However MOSES 1 does not include carbon cycle processes or interactive vegetation, which are both important elements of a comprehensive Earth System model. In order to include these features, the newer MOSES 2.2 model (Essery et al., 2003) has been incorporated into FAMOUS. MOSES2.2 describes the fluxes of CO<sub>2</sub>, water, heat and momentum at the interface between the

---

## 1 Model description and motivation

The climate model used in this work is FAMOUS (Jones et al., 2005; Smith et al., 2008), which is a lower resolution version of the HadCM3 climate model (Pope et al., 2000; Gordon et al., 2000). The atmospheric component

land and the atmospheric boundary layer, and is capable of hosting a number of sub-gridscale tiles in each gridbox, allowing a degree of heterogeneity in surface characteristics to be modelled.

MOSES 2.2 can function in two modes, either calculating surface exchange fluxes for each surface type individually and then averaging them into a gridbox mean for the atmosphere model, or by aggregating the characteristics of the different surface types together before calculating a single, common exchange flux for the gridbox. The latter mode is used in this work, as it has been found to produce better results in early tests of MOSES2.2 in FAMOUS. It is possible to run MOSES2.2 using static or dynamic vegetation, the latter using the TRIFFID dynamic vegetation model (Cox, 2001).

Subgrid land surface processes present in the simulations presented here are due to five different plant functional types (PFTs) as represented by the TRIFFID dynamic vegetation model; broadleaf trees (BT), needleleaf trees (NT), C3 and C4 vegetation and shrubs. In addition to these PFTs, MOSES 2.2 also calculates fluxes due to four non-vegetation surface types; urban environments, inland water, bare soil and land ice (which is constrained to a gridbox coverage fraction of either 0.0 or 1.0 only). Future research with this configuration of FAMOUS will, in part, aim to examine climates of the past where human intervention in the structure of the land surface was negligible or zero. Therefore, the urban fraction is set to zero throughout this work.

The TRIFFID model dynamically updates the five PFT distributions (and soil carbon content) using a combination of “carbon balance” and inter-PFT competition (Cox, 2001). The carbon balance is itself derived from the MOSES surface exchange scheme (see above). Other PFT-dependent limiting factors affecting plant growth within the TRIFFID model include the presence or absence of light (and its subsequent effect on photosynthesis) and photosynthetic enzymes (Cox, 2001). The combination of the “2.2” version of MOSES with TRIFFID had not been used before in FAMOUS and this was the reason for the perturbed physics ensemble presented here.

In addition to land surface processes, the ocean carbon cycle is also simulated within the model. This sub-model is known as HadOCC, the Hadley Centre Ocean Carbon Cycle model (Palmer and Totterdell, 2001). HadOCC is an “ecosystem model” due to its explicit inclusion of phytoplankton and zooplankton populations. Phytoplankton productivity in the model is limited by the availability of nitrate (the only nutrient simulated) and light and therefore primary production below the photic zone of the ocean is greatly reduced. In addition to plankton, total CO<sub>2</sub>, alkalinity and detrital material densities are calculated through a system of coupled differential equations describing, for example, zooplankton grazing and detrital sinking due to gravity. The interested reader is referred to the Appendix of Palmer and Totterdell (2001) for a full description of the equation system. The carbon cycle model is fully coupled to the physical ocean

model and hence all compartments are subject to advective and diffusive transport. It is important to note that HadOCC is purely biological in nature, meaning that nitrate (strictly nitrate plus ammonium) is not lost or gained due to sedimentation or due to addition from rivers for example. The flux of carbon through the NPZD (nutrient-phytoplankton-zooplankton-detritus) model is coupled to the prognostic flux of nitrogen through constant C:N, “Redfield”, ratios (Palmer and Totterdell, 2001; Redfield, 1958).

Climate models contain many adjustable parameters, each with an associated uncertainty. This uncertainty comes, for some parameters, from the inability to measure the value of an observable to arbitrary accuracy. For example  $N_{L0}$  – the ratio of nitrogen to carbon in a leaf, a model constant representative of a given plant functional type (e.g. shrubs) – is a measurable quantity at the plant leaf scale. The uncertainty associated with this parameter comes from upscaling site measurements to a global quantity. There is also some uncertainty from structural parameters in model parameterisations, which do not have a directly observable equivalent in the real world. For example, LAI<sub>min</sub> is a competition parameter which controls how plants will expand. This is not a directly observable quantity, instead the plausible uncertainty ranges are established largely from insight from the model developers based on how variations in this parameter influence properties of the simulations that are observable, such as forest extent. Previous versions of FAMOUS have had their parameters tuned through different procedures (Jones et al., 2005; Smith et al., 2008; Gregoire et al., 2010), but the combination of a complex land surface scheme coupled to dynamic vegetation and an ocean carbon cycle has not been used before in FAMOUS. The computational efficiency of FAMOUS provides an opportunity to explore relationships between parameters and model response and hence identify the set of structural parameters in this new model which give the highest fidelity output when compared to appropriate observations. To this end, building on the tuning of atmosphere and ocean parameters by Gregoire et al. (2010), two 100 member perturbed physics ensembles were performed; one for the land surface and one for the ocean carbon cycle variables. The full coupling of the terrestrial and ocean carbon cycles is ongoing and will be described in a forthcoming paper.

For both the land surface and the ocean perturbed physics ensembles, the set up of the control run was the same. Constant, preindustrial levels of CO<sub>2</sub> in the atmosphere (290 ppmv) were used. For all simulations using dynamic vegetation in this study, an accelerated mode was used, which enables more rapid convergence of the final distribution of PFTs under constant forcing conditions. This works by coupling the vegetation scheme to the surface exchange scheme only every 5 yr (although this time period can be altered if desired) and thereby exchanging the carbon flux output during that time with the vegetation. After each iteration of this coupling, the dynamic vegetation model is then run asyn-

chronously using a large timestep of 100 yr. This enables equilibrated states of even the slowest responding variables to be approached more rapidly. More information on the technical details of this coupling can be found in Cox (2001). For the vegetation distribution, the control and all ensemble members were initialised at 1860 values and each ensemble member was run for 200 yr of physical-climate time; this was found to be more than sufficient for equilibrium to be reached, particularly bearing in mind that the dynamic vegetation model is run in an accelerated fashion.

For the ocean ensemble, a run length of 200 yr was also found to be sufficient for the variables of interest to equilibrate, even when the ocean tracers were initialised with constant values throughout the ocean. It should be noted that there is no equivalent accelerated mode for the ocean carbon cycle as is used for the land surface. To bring the deep ocean into thermal and carbon equilibrium with the surface would take several thousand years and so it is unfeasible to run a 100 member ensemble where each member is run for this long. The ocean ensemble is validated using near-surface observations (5 m depth) where equilibrium is easily reached in 200 yr. Climatologies were constructed for the last 30 yr of each ensemble member for both the land surface and ocean ensembles.

## 2 Perturbed parameters – land surface

The number of structural parameters present in this version of FAMOUS is large and since the main departure from previous versions concerns the carbon cycle (both on land and in the ocean) it was deemed appropriate to find an optimum set of parameters which best reflect the present day status of the biosphere.

Previous work (Booth et al., 2012) used the Latin hypercube method (e.g. Gregoire et al., 2010) to efficiently sample parameter space within bounds reflecting the uncertainty with which these model parameters are known. Booth et al. were then able to demonstrate that uncertainties in the values of carbon cycle parameters can give rise to significant uncertainty in projections of future climate. The present study also uses the Latin hypercube method to vary the same parameters as Booth et al. [over the same ranges of values](#). (with the addition of  $R_{\text{grow}}$ ) which are described in Table 1. Note that the values for all the plant functional types (PFTs) are co-varied, i.e., if the value of certain parameter for broadleaf trees is doubled, the equivalent parameters for the 4 other PFTs will also be doubled, as in Booth et al. (2012).

The parameters in Table 1 are now described in detail.

- $N_{\text{L0}}$  – The “top leaf nitrogen concentration”. This is defined as the amount of nitrogen per amount of carbon and has the units kg N/kg C (Cox et al., 1999).
- $f_0$  – The ratio of  $\text{CO}_2$  concentrations inside and outside leaves at zero humidity deficit (Cox et al., 1999).

- $\text{LAI}_{\text{min}}$  – Any PFT must achieve this value of the Leaf Area Index before it starts to contend with other PFTs for growing area (Cox, 2001).
- $Q_{10}$  – This parameter describes the how the respiration rate of soil varies with temperature. This is done using a power law multiplier, the exponent of which rises by 1.0 when the temperature rises by  $10^\circ\text{C}$  (Cox et al., 1999).
  - The “KAPS” parameter, which describes the specific rate of soil respiration at  $25^\circ\text{C}$  and at optimal soil moisture, is co-varied with  $Q_{10}$  to maintain respiration at this temperature at the standard model rate.
- $V_{\text{crit},\alpha}$  – This is a new parameter which has been integrated into the model code and is defined by  $V_{\text{crit}} = V_{\text{wilt}} + V_{\text{crit},\alpha} (V_{\text{sat}} - V_{\text{wilt}})$  where,  $V_{\text{crit}}$ ,  $V_{\text{sat}}$  and  $V_{\text{wilt}}$  are “by volume” soil moisture concentrations ( $\text{m}^3$  of water per  $\text{m}^3$  of soil). Below  $V_{\text{wilt}}$ , leaf stomata close;  $V_{\text{sat}}$  is the soil moisture amount at the point of saturation and  $V_{\text{crit}}$  is the amount above which PFTs are not water limited. The fact that  $V_{\text{crit},\alpha}$  varies between zero and one means that  $V_{\text{crit}}$  varies between  $V_{\text{wilt}}$  and  $V_{\text{sat}}$  (Cox et al., 1999).
- $T_{\text{upp}}$  – This is one of two parameters which affect how photosynthesis varies with temperature (Cox et al., 2000), the other being  $T_{\text{low}}$ . As can be seen from Table 1, there is actually only one free parameter for  $T_{\text{upp}}$ , because the values for NT, C3, C4 and shrubs are also co-varied. In addition, the values of  $T_{\text{low}}$  are as follows;  $T_{\text{low,BT}} = T_{\text{upp,BT}} - 36$ ,  $T_{\text{low,NT}} = T_{\text{upp,BT}} - 41$ ,  $T_{\text{low,C3}} = T_{\text{upp,BT}} - 36$ ,  $T_{\text{low,C4}} = T_{\text{upp,BT}} - 23$ ,  $T_{\text{low,shrub}} = T_{\text{upp,BT}} - 36$ .
  - Booth et al. (2012) present a variable transformation and define  $T_{\text{opt}} = T_{\text{upp}} - 4.0$  here. This is because  $T_{\text{opt}}$  is more directly observable. The full definitions of  $T_{\text{upp}}$  and  $T_{\text{low}}$  are retained here for completeness and to aid the understanding of the model user.
- $R_{\text{grow}}$  – The “growth respiration fraction”. The total respiration,  $R_{\text{p}}$ , of plants can be divided into those amounts required for the maintenance,  $R_{\text{pm}}$ , and growth,  $R_{\text{pg}}$ , of the plant, where  $R_{\text{pg}}$  is defined as  $R_{\text{pg}} = R_{\text{grow}} (\Pi_{\text{G}} - R_{\text{pm}})$ , and  $\Pi_{\text{G}}$  is the “gross canopy photosynthesis” (Cox et al., 1999). A corollary of this set of definitions is that  $R_{\text{pg}}$  is also equal to one third of the net primary productivity,  $\Pi = \Pi_{\text{G}} - R_{\text{p}}$ . More information on the precise definition of these parameters can be found in Cox et al. (1999).

Previous work by Gregoire et al. also used an ensemble approach to identify optimal configurations of FAMOUS with respect to atmosphere and ocean parameters which are



known to have a significant effect on the climatology (Gregoire et al., 2010; Jones et al., 2005; Murphy et al., 2004). It was therefore desirable that the results of this earlier work were incorporated into the present optimisation framework and, to this end, the ten highest scoring models from Gregoire et al. were sampled using a further “state parameter”,  $\beta$ . The incorporation of this extra parameter means that it is not just the carbon cycle’s uncertainties which are being perturbed in the ensemble but also those of the physical atmosphere and ocean which have previously been shown have a significant impact on model climate (Jones et al., 2005). The fact that only the 10 highest scoring models from Gregoire et al. are chosen for examination here means that it is only the more plausible combinations of values of the physical parameters which are sampled.

The state parameter,  $\beta$ , was varied continuously between 0 and 1 using the same Latin hypercube sampling technique as for all the other model parameters. However,  $\beta$  was then converted to an integer value between one and ten which was used to discriminate between the ten highest scoring sets of parameters from Gregoire et al. Therefore, in total, eight free parameters were varied and an ensemble of one hundred members was run. For Latin hypercube sampling, it is advantageous to have at least ten times as many ensemble members as free parameters; this condition is therefore easily fulfilled in this case. It would have been statistically advantageous to vary each parameter independently for each PFT but this would have increased the necessary size of the ensemble beyond that which was possible due to computational constraints.

### 3 Perturbed parameters – ocean

A further ensemble, perturbing the parameters in the HadOCC sub-model was also carried out. Table 2 shows the control values of the structural parameters in the ocean carbon cycle of FAMOUS which are varied in this work (see Table 2 of Palmer and Totterdell (2001) for more detailed information). Since there are twenty structural parameters listed in Table 2, to vary each parameter individually would require at least two hundred simulations to be performed which is currently impractical. Therefore, the parameters were subdivided into five categories by their compartmentalisation in the model (the “free parameter index” in Table 2): (1) C:N ratio (2) phytoplankton-specific parameters (3) zooplankton-specific parameters (4) detritus-specific parameters (5) carbonate precipitation. Each parameter represented by these five indices was co-varied and therefore the condition of having at least ten times as many ensemble members as free model parameters (i.e. 5) is met. This method of co-variation was decided upon after discussions with the HadOCC code developers (Paul Halloran, Met Office Hadley Centre, personal communication, 2011) and is in line with the work of Booth et al. (2012) whose co-variation scheme is used here

for the land surface parameter perturbations. All parameters in Table 2 were varied by  $\pm 50\%$  in the Latin hypercube-generated ensemble and, as with the land surface, an ensemble of 100 members was run.

The method of studying uncertainty by varying ocean biogeochemical parameters by  $\pm 50\%$  has been used before by Kriest et al., (2012) and was the basis for doing so in this study. The study of Scott et al. (2011) also uses a perturbed physics approach to parameter uncertainty in the HadOCC model, although in a 1D sense. This reduction in model dimensionality enables all parameters to be varied independently rather than using a co-variation method as used here. As discussed above, this was not feasible in this study due to computational (that is, time) constraints.

Due to the inclusion of the state parameter,  $\beta$ , in the land carbon cycle simulations, some ocean parameters differ between the best land surface and ocean simulations. It has been shown however that these differences to the ocean diffusivity and viscosity (Gregoire et al., 2010) make no significant difference to the model climatology.

## 4 How the perturbed physics ensembles were evaluated

### 4.1 Land surface

#### 4.1.1 The Amazon now

Evaluation of how well the land surface ensemble members matched observations was done by comparison with data adapted from the Advanced Very High Resolution Radiometer (Loveland et al., 2000). This dataset was constructed via a joint European-American project, coordinated through the International Geosphere – Biosphere Programme. The construction of this dataset utilised advanced quality-control techniques and was the first global database of land-surface cover categories produced from high resolution (1km) resolution satellite data.

Figure 1 shows which of the surface types used in TRIFID has the highest fraction within each grid box and additionally what the fractional coverage of the dominant tile fraction in each gridbox is equal to. From this figure it is clear that there are large areas of the world where the dominant tile fraction is significantly different from 1. The global average of the quantity given in the right-hand side of Fig. 1 is 0.63 and the spatial standard deviation is 0.18. The equivalent value for the ensemble mean is 0.72 with a spatial standard deviation of 0.12. The combination of these values (higher mean, lower variability) show that the simulations tend to favour non-coexisting PFTs in each gridbox, compared to observations. In the discussion to follow, Fig. 2 is shown which is the same as Fig. 1 but for the ensemble member which is identified as having the most suitable set of parameters. This clearly shows that the dominant tile fraction in the simulations is significantly higher than observed (Fig. 1). For this

reason, the dominant PFT in a gridbox is used to evaluate the efficacy of the different ensemble members' reproduction of vegetation cover. Figure 1 shows that the Amazon region is a good one to concentrate on because it is a large area where the observed fraction of the dominant surface type is close to 1 and also because of the region's known effects on global climate (e.g. Werth and Avissar, 2002) and terrestrial carbon budget (e.g. Cox et al., 2000).

The Amazon region is defined to be 40° W–80° W, –20° S–10° N in this work and is predominantly defined by its BT coverage (Fig. 1). In this region there are 28 land gridboxes and in the observations 22 are BT, 4 are C4, 1 is bare soil and 1 is shrub. Figure 3 shows a histogram of the fractional agreement between PFTs in the ensemble and the observations, that is, how many of the 28 grid boxes are assigned the same PFT in the ensemble members and in observations. In this instance the term “PFTs” is broadened to include bare soil cover.

Figure 3 shows that the majority (80) of the ensemble members agree with the observations in less than half of the gridboxes in the Amazon region. Of the remaining 20 members, 9 have 50 %–60 % agreement, 10 have 60 %–70 % agreement and 1 does better than 70 %. To reduce the number of ensemble members for inclusion in the search for a credible set of carbon cycle parameters, the top 10 scoring members are chosen for further investigation, this is done by examining the dominant PFT globally.

Amongst the top 10 scoring simulations, there are some common biases such as the overestimation of the NT density over North America and the C3 fraction over Northern Eurasia. In addition to these, the models do not reproduce the observed NT distribution over Eurasia and, although the distribution is promising, the global density of BT is somewhat overestimated. It should be noted that over large parts of these areas, the fractional coverage of the dominant PFT is approximately 50 % or less in the observations (Fig. 1) whereas in the 10 best ensemble members, the fractional coverage is often well over 70 % and sometimes over 90 %. This highlights a characteristic feature of the PFT density calculations internal to the TRIFFID model; coexisting PFTs are minimised compared to observations.

Of the top 10 models, a further 3 are discarded due to the almost complete coverage of northern Eurasia with C3 vegetation and so in summary, 7 ensemble members (termed the  $\alpha 7$  simulations) are left for further consideration albeit with some common biases in their reproduction of contemporary vegetation cover.

It could be argued that training the perturbed parameter ensemble specifically on the Amazon region will tend to ‘overfit’ the ensemble to the observed broadleaf tree distribution in this region. However, it has been found that ensemble members which give a good reproduction of the observed dominant gridbox PFT in the Amazon also tend to do better globally too. This is illustrated in Figure 4, which shows the dominant gridbox PFT for (a) the observations, (b)

HadCM3, (c–e) the three top performing  $\alpha 7$  simulations and (f–h) three further ensemble members with decreasing fractional agreement over the Amazon region. Firstly considering sub-figures (a–c) in Figure 4, it is apparent that HadCM3 (which is the ‘parent’ model of FAMOUS) performs better than the top performing  $\alpha 7$  ensemble member. For example, the NT distribution over North America and the extent of the bare soil region representing the Sahara Desert are more in line with observations than FAMOUS. These improvements notwithstanding, there is still notable disagreement between observations and HadCM3 in several regions, such as the overestimation of BT in sub-Saharan Africa and the underestimation of C3 vegetation in western North America.

The biases common to the  $\alpha 7$  ensemble members noted above are clearly visible in Figures 4 (c–e) (overestimation of NT in North America and C3 in Northern Eurasia) and the similarity of the global dominant PFT distribution in these figures is striking. This is important because it shows that training the ensemble on the Amazon region does not lead to significant differences in PFT distributions in other areas of the world and hence that the resulting PFT distributions are robust. The range of each of the parameters varied in this ensemble are indicated below in Fig. 5 for the top three  $\alpha 7$  simulations shown in Figure 4. Figures 4 (f–h) show example ensemble members with decreasing observational agreement over the Amazon region. Crucially, the global agreement also decreases markedly so that in the case of Figure 4 (h), the agreement between the simulated and observational PFT distribution is virtually non-existent.

In summary, it has been shown that the agreement between the dominant PFT in perturbed physics ensemble members and that in observations is a good indicator of the fidelity of global PFT reproduction. Globally, notable regional biases remain in the ensemble members with the leading agreement in the Amazon region but two aspects of this remaining disagreement should be borne in mind.

1. The initial reason for choosing the Amazon region as the training area for the ensemble was due to it having a dominant-PFT fractional coverage close to unity (Fig. 1) which is an emergent property of the simulations (Fig. 2 for example). The only other area of any real size with this property is the Sahara, which is by definition covered with bare soil.
2. It is likely that if the ensemble had been trained to a different area of the globe that a different resulting PFT distribution would have resulted. The aim of this study is not *per se* to show that if the Amazon region's plant biosphere can be reproduced accurately then the rest of the world's PFTs will follow suit. It is known however that the Amazon exerts a significant influence on global climate (Werth and Avissar, 2002) and so a model which can reproduce the gross features of its land surface properties was deemed crucial, although admittedly somewhat subjective.

#### 4.1.2 Sensitivity of results to perturbed parameters

The 8 individual free parameters all influence different aspects of the land surface and hence the wider climate response in the model. Selecting the 7 sets of optimal parameter combinations (the  $\alpha 7$  simulations) tells us something about how the observed metrics can constrain these parameter ranges. If the  $\alpha 7$  simulations all correspond to similar values of a certain parameter, then this is an indication that only a relatively small range of the currently considered plausible parameter space is consistent with observed land surface coverage. This is illustrated in Fig. 5 where the individual parameter values plotted on the vertical axis are normalised between 0 and 1, where 0 represents the lowest value of the parameter chosen by the Latin hypercube sampling, and 1 represents the highest, with all other values being linearly interpolated between the two. *Note that the top three simulations (in terms of their fractional agreement with the observed dominant PFT in the Amazon region) are shown with different symbols to the other  $\alpha 7$  ensemble members for clarity. The global PFT agreement for these three simulations is shown in Fig. 1.*

Figure 5 shows that some of the credible parameter ranges obtained from the ensemble are considerably smaller than others. For example,  $T_{\text{upp}}$  could take essentially any value sampled in the ensemble, whereas  $f_0$  is found to be limited to higher values and  $V_{\text{crit},\alpha}$  to lower values. *The methodology was not developed to place formal constraints on individual parameter ranges, nevertheless, Figure 5 does indicate that the comparison of simulated and observed Broadleaf extent could be used to narrow the plausible range for a number of parameters.* Numerically, the parameters are fractionally constrained as follows:  $f_0$  (31 %),  $\text{LAI}_{\text{min}}$  (88 %),  $N_{\text{L0}}$  (53 %),  $R_{\text{grow}}$  (62 %),  $T_{\text{upp}}$  (92 %),  $Q_{10}$  (78 %),  $V_{\text{crit},\alpha}$  (30 %) and  $\beta$  (59 %).

The fact that the largest parameter uncertainty lies with  $T_{\text{upp}}$  poses a challenge for future carbon cycle changes, where temperature dependence of plant photosynthesis (represented by this parameter) is the dominant uncertainty in future responses (Booth et al., 2012). This result suggests that contemporary plant distributions do not provide a potential constraint on the range of plausible  $T_{\text{upp}}$  values, and hence a way to constrain the range of future changes. This analysis, however, does illustrate that model comparisons with observed vegetation cover may provide a stronger constraint on other parameters ( $f_0$ ,  $N_{\text{L0}}$  and  $V_{\text{crit},\alpha}$ ), *which is novel. Indeed, this approach suggests that observed forest cover may provide a metric to reduce uncertainty ranges on these parameters which have important roles in the climate's hydrological response.*

#### 4.1.3 The Amazon in the past

The Amazon rainforest has been part of the landscape of South America for millions of years. However, its structure

has not remained constant throughout that time (Maslin et al., 2005). Since the reproduction of the structure of the Amazon is highly sensitive to model parameters (see Fig. 3 for example), it is important to further validate the model by perturbing the simulations in another way. This is done by changing the orbital forcing of the  $\alpha 7$  simulations. It is known that the forest's structure was similar to today during the Mid-Holocene (6000 yr ago) and so the  $\alpha 7$  simulations were run for an orbital configuration corresponding to 6000 yr ago and compared to the equivalent for the present day. The leaf area index (LAI) is a parameterisation of the area of leaf cover per unit area of ground (Law et al., 2008) and the differences between the mid-Holocene (and LGM) and their  $\alpha 7$  equivalents are shown in Fig. 6.

It is clear from Fig. 6 that the LAI is generally increased across the Amazon for all of the mid-Holocene simulations with the exception of that shown in Fig. 6a. Maslin et al. have also shown that at the Last Glacial Maximum (LGM) 21 000 yr ago, the density of the Amazon was reduced, as represented by a reduction in LAI. Only Fig. 6h shows a considerable reduction in LAI at the LGM, as required for agreement with the work of Maslin et al. and this is in agreement with the result in Fig. 6a which also identifies this simulation as containing a suitable set of parameters. Therefore a combination of present day observations and paleoclimatic reconstructions of the Amazon rainforest has been used to identify a realistic set of terrestrial carbon cycle parameters suitable for use in further research.

Figure 2 shows the dominant PFT in each gridbox and its fractional coverage for the best performing ensemble member identified in the preceding discussion; it is analogous to Fig. 1 which shows the equivalent data for the observations.

The biases common to the  $\alpha 7$  ensemble members (discussed at the end of Sect. 4.1.1) are clearly seen in Fig. 2, as is the tendency for TRIFFID to not have different PFTS cohabiting in the same gridbox. It should be emphasised that some of these biases may be associated with issues within MOSES/TRIFFID but other biases may be associated with problems with the control climate. For instance, FAMOUS has a tendency to make Australia too wet and hence the Australian desert area is underestimated. Unfortunately, TRIFFID cannot be run offline and hence it is not possible to explicitly separate the climate biases from TRIFFID biases.

## 4.2 Ocean

### 4.2.1 Identification of suitable parameters

The fidelity of the ocean carbon cycle is considered by comparing the concentration of the rate-limiting nutrient in the system, nitrate, with global observations from the World Ocean Atlas (Garcia et al., 2006). The annual mean concentration at 5 m depth in the simulations is compared with the average of the surface and 10 m depth values from the observations. The quality of the model fit to the data is cal-

culated using the Arcsine Mielke skill (AMS) score which gives a score of 1 for perfect correlation and  $-1$  for perfect anti-correlation. If a model field bears no resemblance to the observations then the score will be zero. Further information regarding the AMS can be found in Jones et al. (2005) and Watterson (1996), for example. The nitrate data in the World Ocean Atlas data is given on  $1^\circ$  resolution and therefore it must be regridded onto the model grid of  $2.5^\circ \times 3.75^\circ$  before meaningful comparisons can be made.

Of the 100 ensemble members, 4 gave unphysical values for the nitrate concentration in the climatologies; Fig. 7 shows the remaining 96 members' AMS values.

It is important that when a model parameter is varied to find an optimum configuration, the range of values of that parameter give rise to a broad range of model responses. It is apparent from Fig. 7 that this condition is met for nitrate, where the AMS ranges from 0.040 to 0.72 (mean 0.51) with a standard deviation of 0.16. On the contrary, if one compares the sea surface temperature from the model ensemble with observations from Rayner et al. (2003), the standard deviation is just 0.0017 around a mean of 0.85.

It should be noted here that, in reality, the productivity of the Southern Ocean is iron limited (Boyd et al., 2000). Therefore, as a further check of the validity of this method, the same AMS calculations were performed but excluding ocean points south of  $60^\circ$  S. Even with this restriction on the area of study, the parameter set identified as the best in Fig. 7 still provides an AMS score of 0.67, compared to a maximum of 0.72 and a minimum of 0.05. The average difference between the AMS scores for the global and no-Southern-Ocean cases is  $+0.02$  and the standard deviation of this quantity is 0.04. Therefore the  $+0.05$  difference between the value of 0.72 for the global case and 0.67 for the no-Southern-Ocean case is within this range of variability. It is reassuring that, even excluding the Southern Ocean from the data analysis, the parameters found to give the best global nitrate concentration still give a high fidelity reproduction compared to the majority of the other ensemble members.

The parameters from the highest scoring member of the ensemble (as identified in Fig. 7) are given in Table 3 along with their relationship to the control value. It is encouraging that all but one of the 5 free model parameters deviate noticeably from the control value as it adds weight to the necessity of the exercise. Additionally, none of the 5 parameters are at the extremes of the distribution of parameter space ( $\pm 50\%$ ) when compared to the control simulation, which shows that the postulated range of parameters is plausible.

In addition, Doney et al. (2004) have shown that the background physical state (e.g. the ocean circulation) is perhaps more important for the realism of the ocean carbon cycle than the model parameters themselves. These studies, along with the comparison to observed ocean nitrate concentration performed here, clearly show that a more coordinated study of ocean carbon cycle parameter uncertainties is required and that the work presented here is a step towards achieving the

goal of better constrained parameters affecting the global carbon budget.

Although the ensemble member with the highest AMS is clear from Fig. 7, there are several other ensemble member whose scores are very similar. In light of this, the range of each parameter (with respect to the control value) for the top six highest scoring ensemble members are given in the right-hand column of Table 3 and plotted in Fig. 8. This figure shows that the free parameters affecting the value of the zooplankton, detritus and carbonate precipitation parameters are poorly constrained, with their values representing the top six scoring ensemble members covering virtually the entire parameter range. However, the C:N ratios are confined to approximately the top two-thirds of explored parameter space and the phytoplankton parameters are constrained to roughly the top 25%. This is precisely analogous to the result for the land surface ensemble which showed that the parameter  $f_0$  is constrained to high values within its perturbed range by fitting to the dominant PFT found in the Amazon region.

The reason for this initial 'singling out' of the six top performing ensemble members is that they were thought to be essentially indistinguishable in Fig. 7. However, for the avoidance of doubt, this test was increased to encompass the top ten simulations. Under these less stringent conditions, the ranges of the C:N ratios and phytoplankton parameters are essentially unchanged compared to the consideration of the top six and show that phytoplankton-specific parameters are well constrained by this work.

As a demonstration of the degree of variability present in the ensemble's reproduction of ocean nitrate, Fig. 9 shows the nitrate concentration for the observations, the top scoring ensemble member and four other ensemble members with decreasing AMS scores. The resulting five simulated nitrate distributions vary hugely in their ability to reproduce observations and the behaviour is not 'linear' with Arcsine Mielke score. For example Fig. 9(d) has an AMS of 0.32 and shows a large positive bias globally, whereas Fig. 9(e) (AMS=0.27) does a good job over the tropics and extra-tropics but has a very large negative biases over high latitudes.

#### 4.2.2 Physiological justification of suitable parameters

Now that a set of suitable parameters has been found to reproduce the observed surface nitrate concentration, it is important to be able to justify their magnitude physiologically. As previously stated, each free parameter was varied by  $\pm 50\%$  around the control value, that is, the value 'hard-wired' into the model code and parameters relating specifically to each of C:N ratios, phytoplankton, zooplankton, detritus and carbonate precipitation were co-varied. This latter approach was designed purely to reduce the number of free parameters in the model from the original number (20) to a number which could be sufficiently sampled using a 100 member ensemble (i.e.  $\lesssim 10$ ).

As shown above, the only parameters which are notably constrained by the six top scoring ensemble members are the parameters relating specifically to phytoplankton which have been shown to lie in approximately the top 25 % of the sampled range (Table 3, Fig. 8). However, even in the top six performing ensemble members, the parameters relating to zooplankton, detritus and carbonate precipitation are barely constrained at all within their  $\pm 50\%$  ranges around the control value. What this means is that although the resulting nitrate concentrations for the top six ensemble members are all virtually indistinguishable (from the perspective of their AMS scores) the individual values of the parameters varied can vary significantly. It should be noted that since all zooplankton and detritus variables in Table 2 are co-varied, it is likely that the parameter set found to give the best AMS overall will contain individual parameters which are outside the plausible range which would have been found had each parameter been varied individually. Indeed, the goal of this study was to identify a suitable parameter set for ocean carbon cycle–climate studies rather than to explicitly elucidate the magnitude of each constituent parameter. In order to achieve this, every individual parameter should ideally be varied independently and this was not computationally feasible within the timescale of this study. In spite of this, the values of each parameter are now examined in relation to literature derived values.

As mentioned above Scott et al. (2011) performed a similar perturbed physics ensemble of HadOCC runs to that carried out here and although they use a considerably larger parameter set than the present authors (1000 sets of parameters), the simulations are run in 1 dimension and for run lengths of just 9 yr to examine the model’s internal sensitivity to model parameters, without calling for model–data comparison as performed here. However, this study does provide a useful cross-validation of the validity of the parameter values identified in this work. Table 3 in the Scott paper shows the ranges of the parameters varied. It is important to note here that not only do Scott et al. not vary exactly the same parameter set as the set perturbed here, they also use some different nomenclature to the original Palmer and Totterdell (2001) paper. Table 4 provides a guide to the names of the parameters varied in this study in relation to those given in Palmer and Totterdell (2001) and in Scott et al. (2011).

Of the parameters which are common to both ensemble approaches (Table 4), only two parameter values identified in this study fall outside the range given by Scott et al. (note that the deep detritus remineralisation rate is neglected here since the range in Scott et al. is derived from its shallow equivalent, not from the literature); these are the linear zooplankton mortality rate,  $\mu_1$ , and the ‘rain ratio’. The value of the former derived here is 0.02594 which falls just outside the literature-derived range of 0.03 – 0.2, however the value found here is indistinguishable to the lower bound of this range when quoted to the same level of precision. The

value of the rain ratio is identified as 0.009729 which lies just outside the range quoted in Scott et al., i.e., 0.013 – 0.25.

Now that plausible parameters have been identified for the land and ocean carbon cycles, it is necessary to examine the climatology of this new version of FAMOUS to ensure that the results obtained do indeed represent an improvement in model skill.

## 5 Climatology and validation

Since the first FAMOUS documentation paper (Jones et al., 2005), there have been a number of improvements made. Smith et al. (2008) described advances in the representation of sea ice and ozone as well as the introduction of the HadOCC ocean carbon cycle component. Smith (2012) shows improved upper level winds through the introduction of a Rayleigh friction term at the top of the atmosphere and also described other changes relating to, for example, ocean–solar radiation interactions and the effect of snow at coastal points due to the fractional land–sea mask in FAMOUS (e.g. Smith et al., 2008). The climatologies of runs using the newly identified carbon cycle parameter sets are now described.

### 5.1 Atmosphere and land surface

#### 5.1.1 Near-surface air temperature

It is important to confirm that the new versions of FAMOUS described here are compatible with those published previously (Jones et al., 2005; Smith et al., 2008; Smith, 2012) and with HadCM3. This is because FAMOUS was originally calibrated against HadCM3 in order to provide an analogous climatology but with significantly reduced run times. The FAMOUS simulations in question (denoted by their unique 5 letter Met Office Unified Model simulation index) are given below and are denoted a generation number to indicate the order of their documentation date. The version of the land surface scheme, MOSES, is also given.

- ADTAN (Jones et al., 2005) – MOSES 1
  - Generation 1
- XDBUA (Smith et al., 2008) – MOSES 1
  - Generation 2
- XFHCC (Smith, 2012) – MOSES 1
  - Generation 3
- XFHCU (optimised carbon cycle parameters) – MOSES 2.2 (fixed vegetation)
  - Generation 4a
- XFHCS (optimised carbon cycle parameters) – MOSES 2.2 + TRIFFID (dynamic vegetation)

#### – Generation 4b

The generation 3 simulation, XFHCC, is the most recently documented version of FAMOUS prior to this work although most work currently being undertaken with FAMOUS uses XDBUA (the generation 2 model) or XFXWB (Smith, 2012). The only major structural difference between XFXWB and XFHCC (the generation 3 model used here) is the inclusion of Rayleigh Friction in the upper 3 atmospheric model levels; a change which has been shown to improve the climatology. XFHCC is therefore chosen above XFXWB as the generation 3 model to enhance traceability in the documentation of FAMOUS; noteworthy differences between XFXWB and XFHCC are described in Smith (2012).

All previously documented versions of FAMOUS have used the MOSES 1 land surface scheme and a fixed vegetation distribution and so the newly optimised description of the model represents a step change in model complexity. Figure 10 shows the 1.5 m air temperature for the simulations described above and Table 5 shows the corresponding AMS values.

One particularly apparent aspect of the FAMOUS results shown in Fig. 10 is the persistent cold bias in the Northern Hemisphere in DJF although this is significantly improved in more recent versions of the model compared to the 1st generation. Generations 3, 4a and 4b are strikingly similar in DJF with a cold bias which is shifted east compared to generation 2. In addition, the agreement between FAMOUS and HadCM3 is noticeably better in JJA compared to DJF; this is evident in all versions of the model.

Another result (not shown) is that the introduction of MOSES 2.2 (with fixed vegetation cover) whilst maintaining the un-optimised carbon cycle parameters overcompensates for the Northern Hemisphere winter cold bias and introduces a summer warm bias. Using the optimised parameter set does leave some cold bias in place (Fig. 10) but significantly improves this “new” summer warm bias. So in summary, the introduction of MOSES 2.2 provides an annual mean temperature climatology which is as good as any of the previously documented versions of FAMOUS. If the vegetation is fixed to observations of the contemporary biosphere, the optimisation procedure described above provides not only a good global AMS score, but also helps to alleviate the persistent DJF Northern Hemisphere cold bias.

#### 5.1.2 Vertical temperature profile

Having studied the ability of FAMOUS to reproduce HadCM3’s surface temperature distribution, the air temperature aloft is now examined with respect to the ECMWF 40 yr reanalysis (Uppala et al., 2005). The vertical temperature structure of FAMOUS was last studied in Smith et al., 2008 (their Figure 5) and Fig. 11 shows an updated version of this figure but with simulation output plotted with respect to ERA-40 data rather than ERA-15.

The atmospheric resolution of FAMOUS is significantly reduced compared to HadCM3 (11 vertical levels compared to 19); indeed there is frequently just a single model layer at pressures lower than the tropopause (Smith et al., 2008). Therefore, the ability of the generation 4b version of FAMOUS to accurately reproduce the temperature structure of HadCM3 and ERA-40 up to 10 mbar (the lowest value pressure level available for all the simulations presented here) is very encouraging.

One reason for the improvement in upper-atmosphere temperature profiles (along with improved upper level winds as described in Smith, 2012) is due to the different ozone parameterisations in the separate model versions and these values are shown in Table 6.

#### 5.1.3 Precipitation

Smith et al. (2008) used the CPC Merged Analysis of Precipitation (CMAP) dataset (Xie and Arkin, 1997) to validate the 2nd generation FAMOUS model, XDBUA, and this dataset is also used here. Figure 12 shows the annual mean total precipitation for the CMAP climatology, the 3rd generation model, XFHCC and the 4th generation models XFHCU (fixed vegetation) and XFHCS (dynamic vegetation). This figure also shows the respective AMS values.

The land surface scheme of a climate model can be expected to have a significant effect on precipitation over land. For example, a significant difference between the land surface schemes of the 4th generation versions of FAMOUS and those documented previously is the introduction of plant functional types which can individually affect the fluxes of water and CO<sub>2</sub> at the land-atmosphere interface.

In light of this, it is reassuring that in both 4th generation versions of FAMOUS the global representation of precipitation is improved compared to the 3rd generation version as shown by the AMS scores in the subtitles to Fig. 12b–d. The main features to note are the improvement to the (positive and negative) biases over the equatorial Pacific in Fig. 12c, d and also over the Amazon basin in Fig. 12c, where the vegetation is held constant. **The significant improvement to the positive bias over the Maritime Continent is also rather striking, particularly given this region’s significant influence on large-scale heating and atmospheric circulation (Neale and Slingo, 2003).** In tandem with these improvements, There is a small increase in the positive bias in the equatorial Atlantic in the 4th generation models but overall the global precipitation is noticeably improved with respect to the earlier version.

Figure 12a clearly shows that the areas of highest rainfall are located in the ITCZ and SPCZ (Inter-Tropical and South Pacific Convergence Zones). What this means is that the precipitation anomalies with respect to the CMAP observations in Fig. 12b–d mainly highlight these areas. Figure 13 shows the same data as Fig. 12 but only for the northern and south-

ern mid-latitudes ( $30^{\circ}$ – $60^{\circ}$ ) and Table 7 gives the respective AMS scores.

From Figs. 12 and 13 as well as Table 7 it can be seen that although the global, tropical and southern mid-latitude AMS is improved in the generation 4 simulations compared to the generation 3 version, this is not the case for northern mid-latitudes. This slight deterioration is due to an increase in the positive bias over Western North America and an increase in the negative bias over the Northern Pacific.

## 5.2 Ocean

### 5.2.1 Surface nitrate

Figure 14 shows the annual average surface nitrate concentration for observations from the World Ocean Atlas (Garcia et al., 2006) and for the generation 3 and 4a simulations.

There are no significant differences between the nitrate distributions for the two generation 4 models XFHCU – Fig. 14c – and XFHCS (not shown) which is expected because these simulations differ only in their representation of terrestrial vegetation. When comparing Fig. 14b and c however, a marked improvement in FAMOUS' ability to reproduce the observed nitrate concentration is seen between generation 3 and 4, which is clearly manifested in a significant increase in the AMS score for XFHCU as shown above Fig. 14b, c. Clearly this is the expected result because the optimised ocean carbon cycle parameters used in XFHCU were tuned to the nitrate concentration in Fig. 14a. However this does provide a good illustration of the power of the tuning method employed in this work. For example, the large positive bias in the equatorial Pacific is significantly reduced and, although the positive bias in the south Atlantic is increased, the overall Southern Ocean bias is markedly reduced. As previously mentioned however, the Southern Ocean bias is of lesser importance here since the ocean productivity in this region is, in reality, iron limited (Boyd et al., 2000).

### 5.2.2 Vertical nitrate profiles

Although the surface nitrate distribution has been improved by the newly identified set of ocean carbon cycle parameters, the effect on the same quantity at depth is now investigated, again with respect to the World Ocean Atlas dataset. Figure 15 shows the observed quantity and the equivalent plots for the generation 3 and 4 configurations of FAMOUS respectively. Although there is a deterioration in the negative bias around 1000m depth in northern mid-latitudes, there is a striking improvement in the positive bias at high northern latitudes. The alleviation of the Southern Ocean negative bias present in the third generation model (shown above in Figure 14) is also significantly improved.

In addition to the global results shown in Figure 15, Figures 16 and 17 show the results for the Atlantic ( $-70^{\circ}$ W –  $20^{\circ}$ E) and Pacific ( $150^{\circ}$ E –  $290^{\circ}$ E) basins respectively. In the

Atlantic, the agreement at depth in the southern hemisphere is less good in the fourth generation models (sub-figures (c) and (d)), however the northern hemisphere agreement is improved at all depths. For the Pacific Ocean, there is little change in the level of agreement in the southern hemisphere except for a small deterioration in the positive bias around 1000m depth between the equator and approximately  $45^{\circ}$ S. The negative bias at depths below 2000m is noticeably improved in the northern hemisphere however.

### 5.2.3 Ocean productivity

Figure 18 shows global and zonally averaged ocean productivity data from observations (Behrenfeld and Falkowski, 1997) and the difference between the generation 3 and 4a configurations of FAMOUS and this observational data (as for the surface nitrate concentration, the generation 4b model is not shown since the results are very similar to those of the generation 4a model). The agreement between simulated and observed distributions is degraded in the equatorial oceans for the generation 4a configuration. However in the extra-tropics, right up to the polar regions, the agreement between the different generations of FAMOUS themselves is striking, although the simulated values are generally less productive than observed. It should be noted that this inter-model agreement in the extra-tropics is not seen for the surface nitrate concentration (Fig. 14) where significant differences are visible in the Southern Ocean for the different model configurations.

The global productivity sum for the observations is  $50.8 \text{ Pg C yr}^{-1}$ , in contrast to  $33.0$  and  $71.3 \text{ Pg C yr}^{-1}$  for generations 3 and 4a respectively. Before drawing conclusions about the efficacy of these results, it must be borne in mind that the simulated figures are for total productivity and therefore it is necessary to subtract off the phytoplankton respiration fraction, as is described in Palmer and Totterdell, 2001. Performing this transformation gives the net productivity of the generation 3 and 4a models at  $26.0$  and  $56.2 \text{ Pg C yr}^{-1}$  respectively, or 51% and 111% of the observed global mean value.

In summary, although there is a degradation in the agreement between observed and simulated results for the generation 4a results compared to generation 3, the behaviour of the models is only substantially different outside the tropics and the global mean value of net productivity, which is often quoted in the literature (Palmer and Totterdell, 2001; Behrenfeld and Falkowski, 1997; Cox et al. 2000, for example) is significantly closer to the observed value for the generation 4a model compared to the previously published generation 3 model.

## 6 Discussions, conclusions and future work

The two new versions of FAMOUS presented here represent an important increase in model complexity compared to previous versions of the model, with the inclusion of surface tiling into 9 sub-types and the flexibility to include dynamic vegetation response to climate forcings. The carbon cycle parameters of both the land surface and the ocean have been tuned to observations and reanalysis data and the climatologies of the new versions of the model have been shown to be noticeably improved.

Concerning the terrestrial carbon cycle, the use of a large ensemble of 100 climate simulations has enabled the determination of sensible ranges of the parameters varied in the ensemble methodology. It is clear that certain parameters are significantly better constrained than others by this work. For example, the 7 ensemble members which are seen to give the best representation of the Amazon rainforest only account for 30 % of the variation of the parameter controlling the critical soil moisture ( $V_{\text{crit},\alpha}$ ) whereas the same 7 simulations encompass 92 % of the parameter range of the  $T_{\text{upp}}$  parameter which, in part, controls the response of photosynthesis with temperature. This last result concerning  $T_{\text{upp}}$  suggests that comparisons with land surface coverage do not provide a constraint on the future land carbon cycle uncertainty identified here and in Booth et al. (2012). It does raise the interesting implication, however, that comparisons of land surface coverage between observations and simulations may constrain other land carbon cycle parameters more closely tied to the hydrological response within the model. For the ocean ensemble, it has been shown that the phytoplankton parameters are better constrained than the others with the top six performing ensemble members accounting for approximately just the top 25 % of the parameter variation.

Despite including many elements of the carbon cycle, the work presented here fixes the atmospheric concentration of  $\text{CO}_2$  at preindustrial levels. This clearly limits the degree to which the newly modelled carbon cycle processes can influence the large-scale climate of the model. Lifting this restriction whilst maintaining a realistic climate simulation, and assessing the climate and sensitivities of this fully interactive carbon cycle version of FAMOUS is beyond the scope of this paper, and will be addressed in a forthcoming publication.

This work illustrates that many parameters are under-constrained or under-determined by the simulation-observation comparisons presented here. This represents one of the key challenges in model development and is an important factor linked to uncertainty in simulated responses to future climate scenarios (e.g. Booth et al., 2012). The use of simulated-observed comparisons to constrain model parameters is more advanced in the development of atmospheric components of models (e.g. Murphy et al., 2004 where simulations are compared against a very large basket of observational metrics) but even in these cases, with the larger number of observations, many (often key)

parameters remain under-determined (Sexton et al., 2012). The development of a comparable set of observational metrics for carbon cycle processes, is in its infancy in comparison. It has been illustrated here that land vegetation cover, specifically Amazon Forest extent, and ocean nitrate can both be used to narrow the range of plausible values for some but not all carbon cycle parameters. This is a first step. With the more central role of carbon cycle processes in current global climate models (CMIP5) we will need to identify and develop a broader set of biogeochemical observational metrics as part of the processes of calibrating parameter sets, while recognising that some parameters will always remain under-determined and that this will be linked to an uncertainty in the simulated responses.

The HadOCC model is currently the only biogeochemical model which can be coupled to the Hadley Centre GCM ocean. Future work with this modelling framework will focus on significantly improving the biogeochemical cycling capabilities of the Hadley Centre model. The aim of the first stage of this biogeochemical cycling improvement work will aim to include oxygen as a fully prognostic variable and later developments will aim to increase the number of nutrients simulated in the model, which is currently limited to just one, i.e. nitrate. More recent versions of the Hadley Centre model (for example HadGEM2 which is part of the Hadley Centre's contribution to the forthcoming IPCC Fifth Assessment Report) use an upgraded version of the HadOCC model known as diat-HadOCC (Halloran, 2012; Collins et al., 2011), however this configuration of the ocean carbon cycle model has not yet been applied to the FAMOUS version of the Hadley Centre model.

Additional future model development work with the FAMOUS biogeochemical scheme aims to incorporate further ocean processes related to long timescale responses of the ocean carbon cycle (such as weathering); analogous to the GENIE Earth system model (e.g. Ridgwell and Hargreaves, 2007). This will enable more realistic multi-centennial climate simulations to be carried out with a view to aiding a better understanding of ocean acidification under climate change, for example. Work is also underway to improve the coupling between FAMOUS and the Glimmer icesheet model by including a detailed representation of sub-gridscale orography and snowpack behaviour.

*Acknowledgement.* JHTW and PJV were supported by funding from Statoil ASA, Norway. JHTW and PJV thank T. L. Leith of Statoil ASA for extensive discussions throughout this work. BBBB was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). The authors thank two anonymous referees for their comments.



## References

- Behrenfeld, M. J., and Falkowski, P. G.: Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.*, 42(1), 1–20, 1997.
- Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., Huntingford, C., Betts, R. A., Harris, G. R., and Lloyd, J.: High sensitivity of future global warming to land carbon cycle processes, *Environ. Res. Lett.*, 7, 024002, doi:10.1088/1748-9326/7/2/024002, 2012.
- Boyd, P. W., Watson, A. J., Law, C. S., Abraham, E. R., Trull, T., Murdoch, R., Bakker, D. C. E., Bowie, A. R., Buesseler, K. O., Chang, H., Charette, M., Croot, P., Downing, K., Frew, R., Gall, M., Hadfield, M., Hall, J., Harvey, M., Jameson, G., LaRoche, J., Liddicoat, M., Ling, R., Maldonado, M. T., McKay, R. M., Nodder, S., Pickmere, S., Pridmore, R., Rintoul, S., Safi, K., Sutton, P., Strzepek, R., Tanneberger, K., Turner, S., Waite A., and Zeldis, J.: A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilisation, *Nature*, 407, 695–702, 2000.
- Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., D, Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A and Woodward, S. Development and evaluation of an Earth-system model–HadGEM2, *Geosci. Model Dev. Discuss*, 4, 1051–1075, doi:10.5194/gmd-4-1051-2011, 2011.
- Cox, P. M.: Description of the “TRIFFID” dynamic global vegetation model, Hadley Centre Technical Note 24, 17 pp., 2001.
- Cox, P. M., Huntingford, C., and Harding, R. J.: A canopy conductance and photosynthesis model for use in a GCM land surface scheme, *J. Hydrol.*, 212, 79–94, 1998.
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM sensitivity of climate and climate sensitivity, *Clim. Dyn.*, 15, 183–203, 1999.
- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Acceleration of global warming due to carbon cycle feedbacks in a coupled climate model, *Nature*, 408, 184–187, 2000.
- Doney, S. C., Lindsay, K., Caldeira, K., Campin, J.-M., Drange, H., Dutay, J.-C., Follows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Madec, G., Maier-Reimer, E., Marshall, J. C., Matear, R. J., Monfray, P., Mouchet, A., Najjar, R., Orr, J. C., Plattner, G.-K., Sarmiento, J., Schlitzer, R., Slater, R., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A.: Evaluating global ocean carbon models: The importance of realistic physics. *Global Biogeochem. Cy.*, 18, GB3017, doi:10.1029/2003GB002150, 2004.
- Eppley, R. W.: Temperature and phytoplankton growth in the sea, *Fish. Bull.*, 70, 1063–1085, 1972.
- Essery, R. L. H., Best, M. J., Betts, R. A., and Cox, P. M.: Explicit representation of subgrid heterogeneity in a GCM land surface scheme, *J. Hydrometeorol.*, 4, 530–543, 2003.
- Garcia, H. E., Locarnini, R. A., Boyer, T. P., and Antonov, J. I.: World Ocean Atlas 2005, Vol. 4: Nutrients (phosphate, nitrate, silicate), edited by: Levitus, S., NOAA Atlas NESDIS 64, US Government Printing Office, Washington, D. C., 396 pp., 2006.
- Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns T. C., Mitchell, J. F. B., and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, 16, 147–168, 2000.
- Gregoire, L. J., Valdes, P. J., Payne, A. J., and Kahana, R.: Optimal tuning of a GCM using modern and glacial constraints, *Clim. Dyn.*, 37, 705–719, doi:10.1007/s00382-010-0934-8, 2010.
- Halloran, P. R.: Does atmospheric CO<sub>2</sub> seasonality play an important role in governing the air-sea flux of CO<sub>2</sub>?, *Biogeosciences*, 9, 2311–2323, doi:10.5194/bg-9-2311-2012, 2012.
- Jones, C., Gregory, J., Thorpe, R., Cox, P., Murphy, J., Sexton, D., and Valdes, P.: Systematic optimisation and climate simulations of FAMOUS, a fast version of HadCM3, *Clim. Dyn.*, 25, 189–204, 2005.
- Kriest, I., Khatiwala, S., and Oschlies, A.: Towards an assessment of simple global marine biogeochemical models of different complexity, *Progr. Oceanogr.* 86, 337–360, doi:10.1016/j.pocean.2010.05.002, 2012.
- Law, B. E., Arkebauer, T., Campbell, J. L., Chen, J., Sun, O., Schwartz, M., van Ingen, C., and Verma, S.: Terrestrial carbon observations: protocols for vegetation sampling and data submission, Food and Agricultural Organisation of the United Nations, Rome, 87 pp., 2008.
- Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *Int. J. Remote Sens.*, 21, 1303–1330, 2000.
- Maslin, M., Malhi, Y., Phillips, O., and Cowling, C.: New views on an old forest: assessing the longevity, resilience and future of the Amazon rainforest, *T. I. Brit. Geogr.*, 30, 477–499, 2005.
- Murphy, J. M., Sexton, D. M. H., Barnett, D., N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768–772, 2004.
- Neale, R., and Slingo, J.: The maritime continent and its role in the global climate: A GCM Study. *J. Clim.*, 16, 834–848, 2003.
- Nishii, K., Miyasaka, T., Nakamura, H., Kosaka, Y., Yokoi, S., Takayabu, Y. N., Endo, H., Ichikawa, H., Oshima, K., Sato, N., and Tsushima, Y.: Relationship of the reproducibility of multiple variables among global climate models. *J. Meteorol. Soc. Japan*, 90A, 87–100, doi:10.2151/jmsj.2012-A04, 2012.
- Palmer, J. R. and Totterdell, I. J.: Production and export in a global ocean ecosystem model, *Deep-Sea Res. Pt. I*, 48, 1169–1198, 2001.
- Pope, V. D., Gallani, M. L., Rowntree P. R., and Stratton, R. A.: The impact of new physical parametrizations in the Hadley Centre climate model – HadAM3, *Clim. Dyn.*, 16, 123–146, 2000.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108, 4407, doi:10.1029/2002JD002670, 2003.
- Redfield, A. C.: The biological control of chemical factors in the environment, *Am. Sci.*, 46, 205–221, 1958.
- Reichler, T., and Kim, J.: How well do coupled models simulate today’s climate? *Bull. Am. Met. Soc.*, 89(3), 303–311, doi:10.1175/BAMS-89-3-303, 2008.
- Ridgwell, A. and Hargreaves, J. C.: Regulation of atmospheric CO<sub>2</sub> by deep-sea sediments in an Earth system model, *Global Biogeochem. Cy.*, 21, GB2008, doi:10.1029/2006GB002764, 2007.

- Scott, V., Kettle, H., and Merchant, C., J.: Sensitivity analysis of an ocean carbon cycle model in the North Atlantic: An investigation of parameters affecting the air-sea CO<sub>2</sub> flux, primary production and export of detritus, *Ocean Sci.*, 7, 405–419, doi:10.5194/os-7-405-2011, 2011.
- Sexton, D. M., Murphy, J. M., Collins, M., and Webb, M. J.: Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Clim. Dyn.*, 38, 2513–2542, doi:10.1007/s00382-011-1208-9, 2012.
- Smith, R. S., Gregory, J. M., and Osprey, A.: A description of the FAMOUS (version XDBUA) climate model and control run. *Geosci. Model Dev.*, 1, 53–68, doi:10.5194/gmd-1-53-2008, 2008.
- Smith, R. S.: The FAMOUS climate model (versions XFXWB and XFHCC): description update to version XDBUA, *Geosci. Model Dev.*, 5, 269–276, doi:10.5194/gmd-5-269-2012, 2012.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, *Q. J. R. Meteorolog. Soc.*, 131, 2961–3012, doi:10.1256/qj.04.176, 2005.
- Watterson, I. G.: Non-dimensional measures of climate model performance, *Int. J. Climatol.*, 16, 379–391, 1996.
- Werth, D., and Avissar, R.: The local and global effects of Amazon deforestation, *J. Geophys. Res.*, 107, 8087, doi:10.1029/2001JD000717, 2002.
- Xie, P., and Arkin, P. A.: Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *B. Am. Meteorol. Soc.*, 78, 2539–2558, 1997.

**Table 1.** List of parameters used in the land surface carbon cycle perturbed physics ensemble. The values of the minimum leaf area index (LAI) for C3, C4 and shrubs are not varied in this work and hence only one value is given. The three different parameters used are (1) the minimum value used in the Latin hypercube sampling scheme (2) the “standard” value used in the simulation framework before parameter perturbation and (3) the maximum value. Note that the ranges used in this work are the same as in Booth et al. (2012). The additional  $R_{\text{grow}}$  parameter in this work is varied by 50 % either side of its standard value.

	BT	NT	C3	C4	Shrub
$N_{\text{Lo}}$	0.018, 0.03, 0.1	0.024, 0.03, 0.082	0.028, 0.06, 0.152	0.018, 0.03, 0.188	0.018, 0.03, 0.096
$f_0$	0.7, 0.875, 0.95	0.7, 0.875, 0.95	0.7, 0.9, 0.95	0.65, 0.8, 0.8	0.7, 0.9, 0.95
$\text{LAI}_{\text{min}}$	1, 3, 4	1, 3, 4	1	1	1
$Q_{10}$	1.5, 2, 3.5	1.5, 2, 3.5	1.5, 2, 3.5	1.5, 2, 3.5	1.5, 2, 3.5
$\alpha$	0, 0.5, 1	0, 0.5, 1	0, 0.5, 1	0, 0.5, 1	0, 0.5, 1
$T_{\text{upp}}$	31, 36, 41	$T_{\text{upp, BT}} - 5.0$	$T_{\text{upp, BT}}$	$T_{\text{upp, BT}} + 9.0$	$T_{\text{upp, BT}}$
$R_{\text{grow}}$	0.125, 0.25, 0.375	0.125, 0.25, 0.375	0.125, 0.25, 0.375	0.125, 0.25, 0.375	0.125, 0.25, 0.375

**Table 2.** Control structural parameters in the HadOCC ecosystem model.

Model parameter name	Control parameter value (units)	Physical interpretation	Free parameter index (see text)
c2n.p	6.625	C:N ratio for phytoplankton	1
c2n.z	5.625	C:N ratio for zooplankton	1
c2n.d	7.5	C:N ratio for detritus	1
psmax	0.6	Maximum rate of photosynthesis	2
alpha	0.02 ((W m <sup>-2</sup> ) <sup>-1</sup> day <sup>-1</sup> )	Initial slope of photosynthesis – irradiance curve	2
Q10H	1.0	Increase in phytoplankton growth rate for a 10 degree temperature increase	2
mort_sat	0.1 (mMol m <sup>-3</sup> )	Half-saturation constant for phytoplankton mortality	2
resp_rate	0.02 (day <sup>-1</sup> )	Rate of phytoplankton respiration in fraction of biomass lost per day	2
pmort_max	0.05 (day <sup>-1</sup> (mMol m <sup>-3</sup> ) <sup>-1</sup> )	Maximum phytoplankton mortality (expressed as biomass fraction lost per day)	2
graze_max	1.0 (day <sup>-1</sup> )	Maximum specific rate of zooplankton grazing	3
graze_sat	0.75 (mMol m <sup>-3</sup> )	Half-saturation constant for zooplankton grazing	3
graze_threshold	0.1 (mMol m <sup>-3</sup> day <sup>-1</sup> )	Threshold for zooplankton grazing function	3
beta_p	0.7	Assimilation efficiency of zooplankton feeding of phytoplankton	3
beta_dt	0.5	Assimilation efficiency of zooplankton feeding on detritus	3
z_mort_1	0.02 (day <sup>-1</sup> )	Linear zooplankton mortality	3
z_mort_2	0.3 (day <sup>-1</sup> (mMol m <sup>-3</sup> ) <sup>-1</sup> )	Quadratic zooplankton mortality	3
remin_rate_shallow	0.1 (day <sup>-1</sup> )	Remineralisation rate, levels 1 to 8	4
remin_rate_deep	0.02 (day <sup>-1</sup> )	Remineralisation rate, levels 9 to 20	4
sink_rate_dt	10.0 (m day <sup>-1</sup> )	Sinking rate for detritus	4
rain_ratio	0.007	Carbon export as calcite, as a proportion of primary production	5

**Table 3.** Parameter values for the highest scoring ocean carbon cycle ensemble member and their relationship to the respective control value. Also shown is the percentage of the control value which is covered by the top six scoring ensemble members, as shown in Figure 8.

Parameter	Values from highest scoring ensemble member	Percentage of control value of highest scoring ensemble member	Range of top 6 scoring ensemble members w.r.t. control
c2n_p	6.559	99 %	91 % – 145 %
c2n_z	5.569		
c2n_d	7.426		
psmax	0.8417	140 %	119 % – 145 %
alpha	0.02806		
Q10H	1.403		
mort_sat	0.1403		
resp_rate	0.02806		
pmort_max	0.07014		
graze_max	1.297	130 %	70 % – 149 %
graze_sat	0.9729		
graze_threshold	0.1297		
beta_p	0.908		
beta_dt	0.6486		
z_mort_1	0.02594		
z_mort_2	0.3891		
remin_rate_shallow	0.06488	65 %	65 % – 139 %
remin_rate_deep	0.01298		
sink_rate_dt	6.488		
rain_ratio	0.009729	139 %	65 % – 140 %

**Table 4.** Comparison of nomenclature used in Scott et al. (2011) and Palmer and Totterdell (2001) to describe analogous ocean biogeochemical parameters.

Model parameter name	Symbol in Table 2 of Palmer and Totterdell (2001)	Symbol in Table 3 of Scott et al. (2011)
c2n_p	$C_p$	not mentioned
c2n_z	$C_z$	not mentioned
c2n_d	$C_d$	not mentioned
psmax	$P_{max10}$	$P_{max}^s$
alpha	$\alpha$	$\alpha$
Q10H	$Q_{10}$	not mentioned
mort_sat	$K_N$	$K_{nit}$
resp_rate	$\eta$	$\eta$
pmort_max	$m_0$	$m_0$
graze_max	$g_{max}$	$g_{max}$
graze_sat	$K_G$	$K_F$
graze_threshold	$F_{threshold}$	not mentioned
beta_p	$\beta_p$	$\beta_P$
beta_dt	$\beta_d$	$\beta_D$
z_mort_1	$\mu_1$	$\mu_1$
z_mort_2	$\mu_2$	$\mu_2$
remin_rate_shallow	$\lambda$	$Rm_{shallow}$
remin_rate_deep	$\lambda$	$Rm_{deep}$
sink_rate_dt	no symbol given	$v_s$
rain_ratio	$\gamma_c$	$\Upsilon_c$

**Table 5.** Regional and seasonal AMS values for different members of the FAMOUS model hierarchy. These are calculated for 1.5 m air temperature with respect to HadCM3. Generation numbers are given in brackets.

		ADTAN (1)	XDBUA (2)	XFHCC (3)	XFHCU (4a)	XFHCS (4b)
45–90° N	DJF	0.60	0.70	0.79	0.82	0.76
	JJA	0.79	0.76	0.84	0.77	0.68
	Annual	0.66	0.77	0.85	0.84	0.76
Global	DJF	0.85	0.86	0.88	0.87	0.87
	JJA	0.90	0.88	0.88	0.88	0.88
	Annual	0.89	0.89	0.89	0.89	0.89

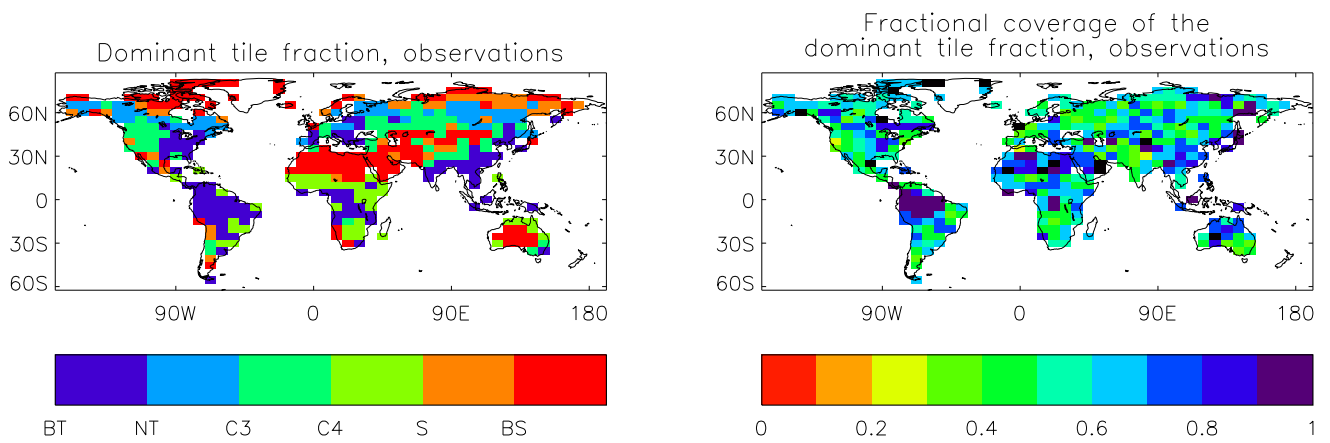
**Table 6.** Ozone concentrations in  $\text{kg kg}^{-1}$  around the tropopause for the different generations of FAMOUS.

Level	1st generation	2nd and 3rd generation	4th generation
Top layer	–	$1.5 \times 10^{-6}$	$6.0 \times 10^{-6}$
Above tropopause	$1.5 \times 10^{-6}$	$1.0 \times 10^{-6}$	$2.0 \times 10^{-6}$
At tropopause	$2.0 \times 10^{-7}$	$2.0 \times 10^{-7}$	$1.0 \times 10^{-7}$
Below tropopause	$2.0 \times 10^{-8}$	$2.0 \times 10^{-8}$	$2.0 \times 10^{-8}$

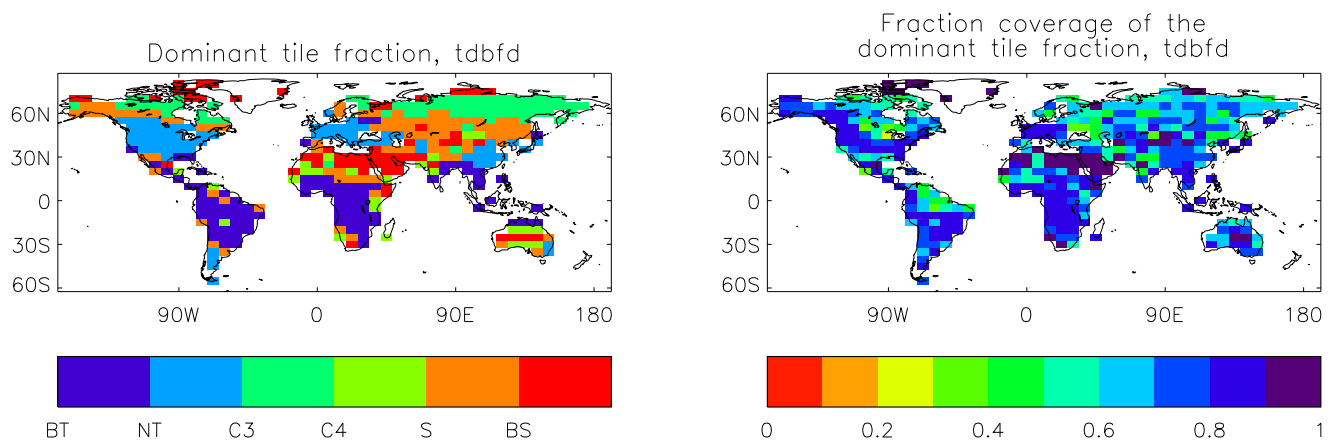


**Table 7.** AMS scores for precipitation for the northern and southern mid-latitudes and the tropics.

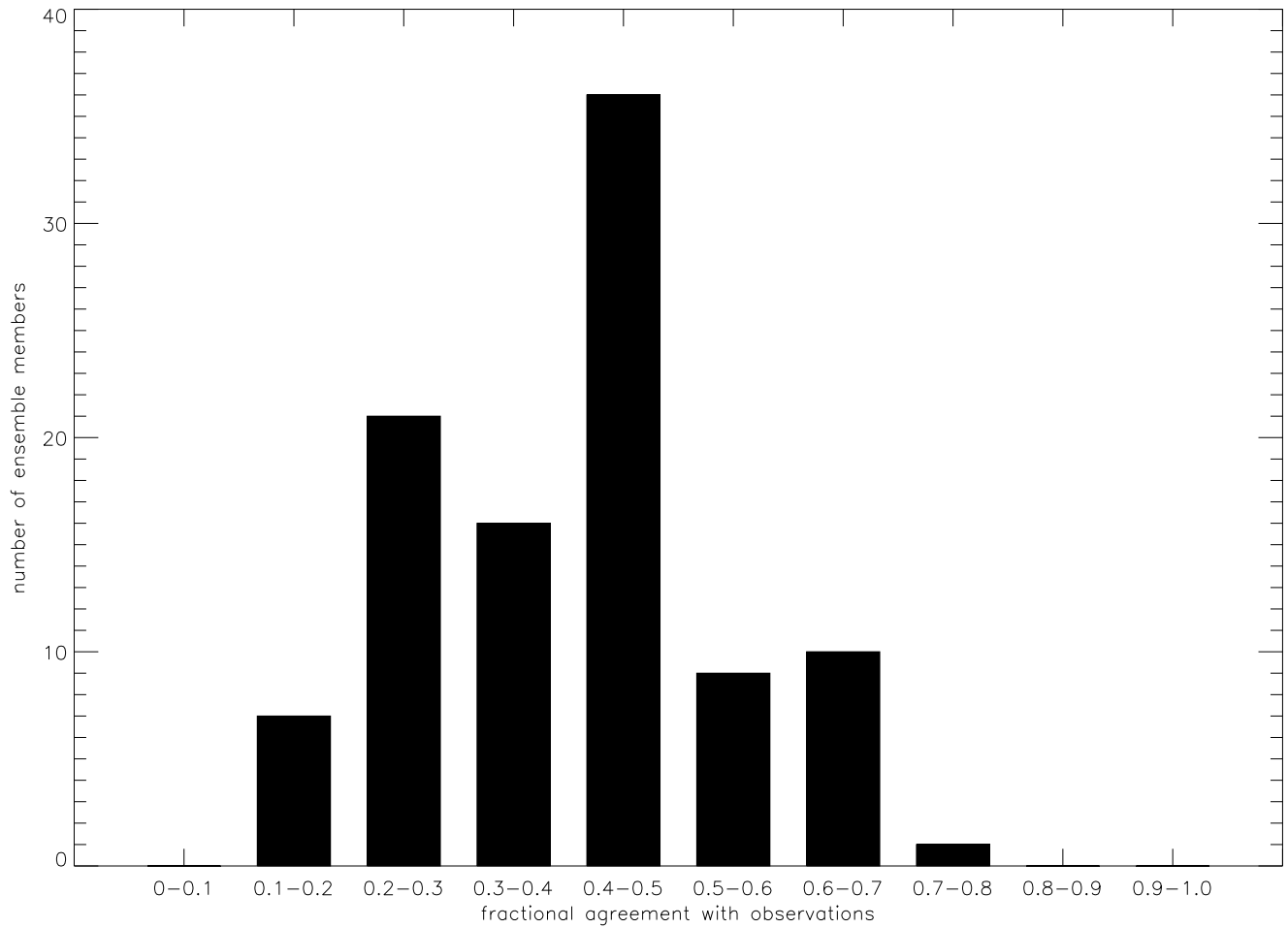
	XFHCC	XFHCU	XFHCS
30°–60°	0.59	0.51	0.51
–30°–+30°	0.40	0.47	0.45
–60°––30°	0.29	0.34	0.36



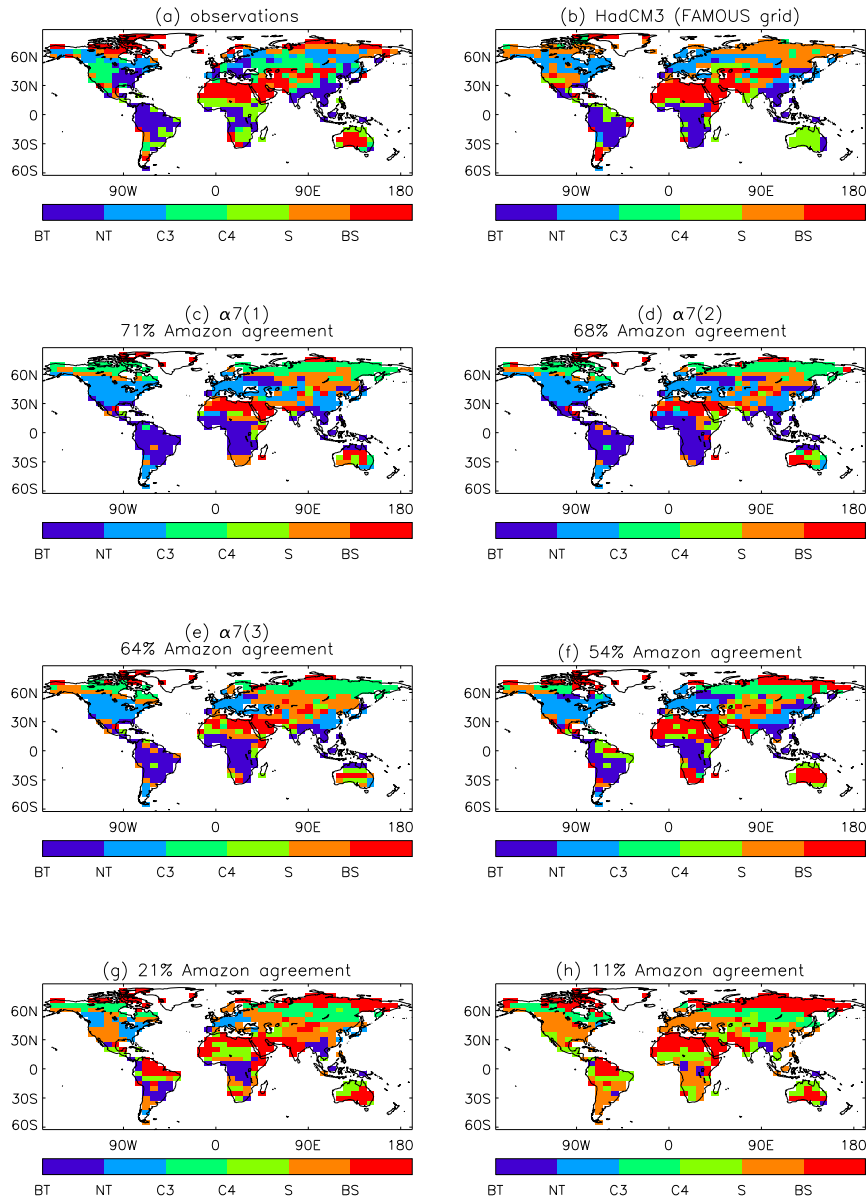
**Fig. 1.** The left-hand figure shows the observed dominant plant functional type for the present day (Loveland et al., 2000) and the right-hand figure shows the fractional coverage of the dominant type. BT (broadleaf tree), NT (needleleaf tree), C3 and C4 vegetation and S (shrubs) and BS (bare soil).



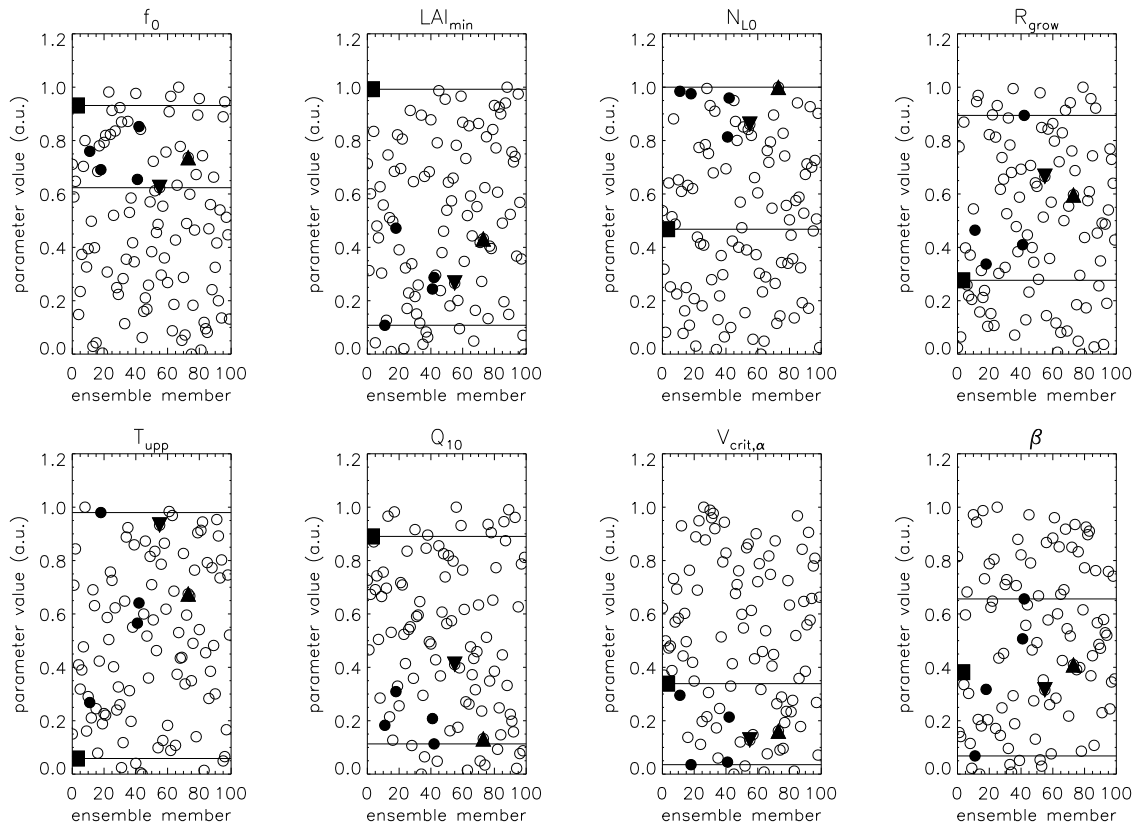
**Fig. 2.** The left-hand figure shows the simulated dominant plant functional type for the best performing land surface ensemble member and the right-hand figure shows the fractional coverage of the dominant type.



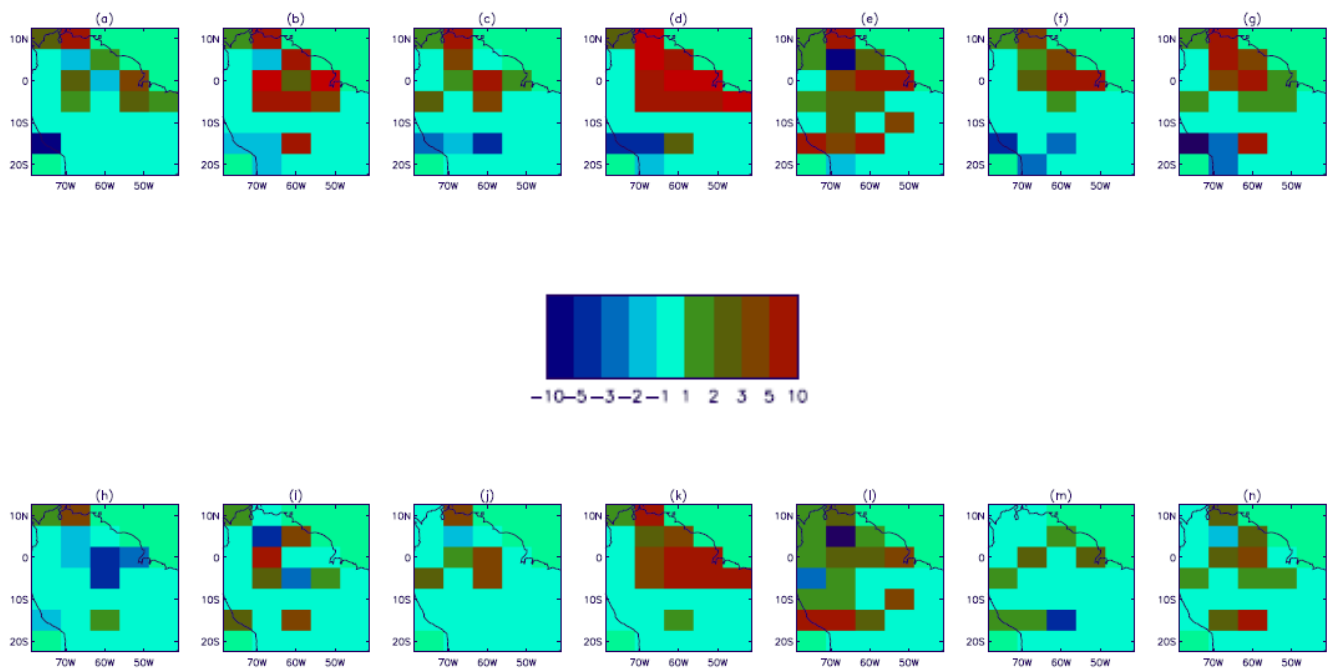
**Fig. 3.** Histogram of the fractional agreement between the 100 ensemble members and the observations over the Amazon region for all PFTs. Here, “fractional agreement”, gives the fraction of the 28 grid Amazonian grid boxes which are assigned the same PFT in the ensemble members and in observations.



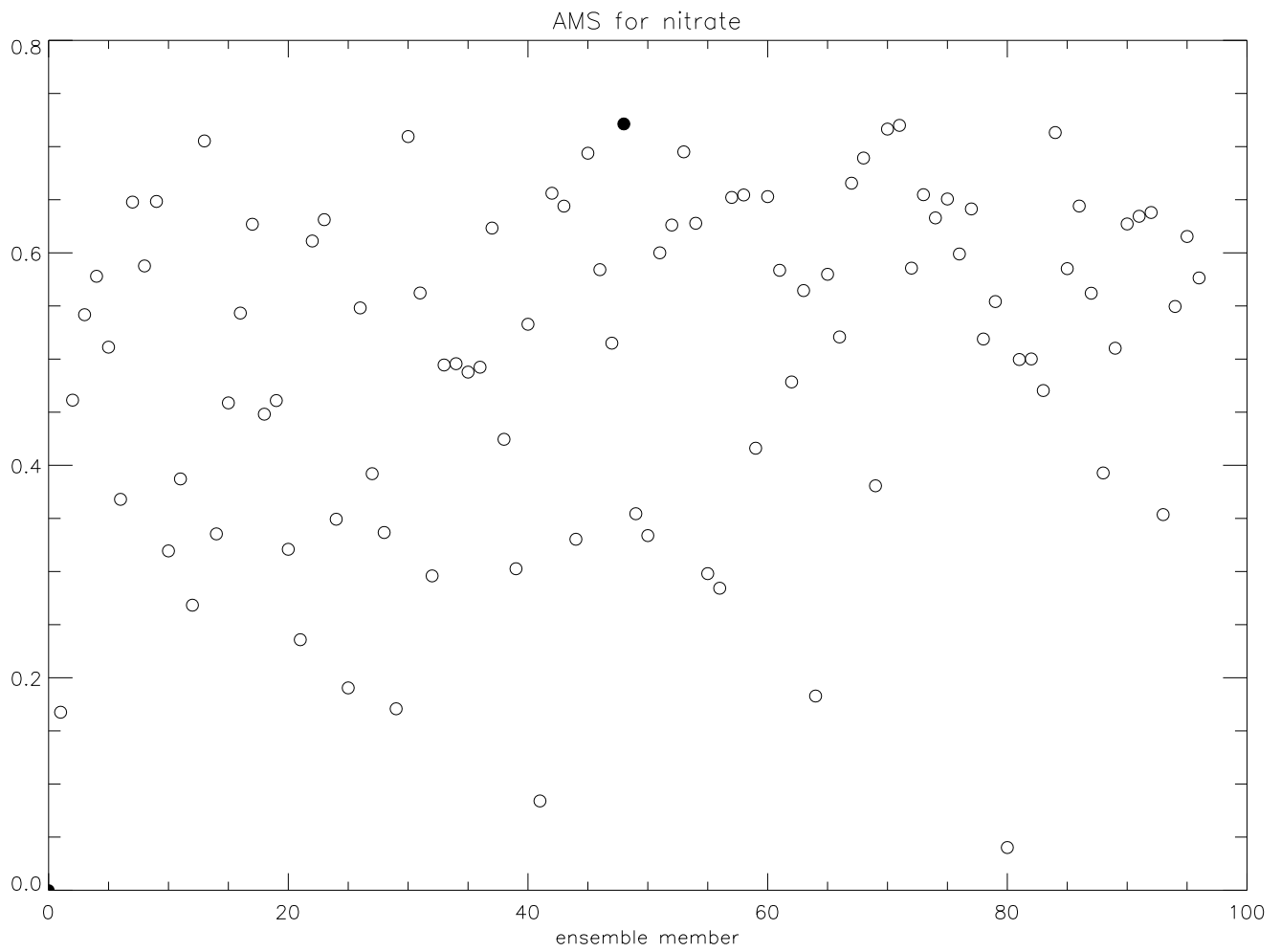
**Fig. 4.** Dominant gridbox PFT for (a) observations, (b) HadCM3, regridded to the FAMOUS resolution (c–e) the three top performing  $\alpha 7$  ensemble members in terms of their fractional agreement with observations over the Amazon region, (f–h) three ensemble members with decreasing Amazon agreement. There is a clear correlation between the agreement over the Amazon region and that over the whole globe.



**Fig. 5.** The sensitivity of the 100 land surface ensemble members to individual parameters. The  $\alpha 7$  simulations are shown with filled symbols and the horizontal lines represent the minimum and maximum values of each parameter covered by them. In decreasing order, the top three performing  $\alpha 7$  simulations (in terms of their fractional agreement with the observed dominant PFT in the Amazon region) are shown by upward-facing arrows, downward-facing arrows and squares respectively.

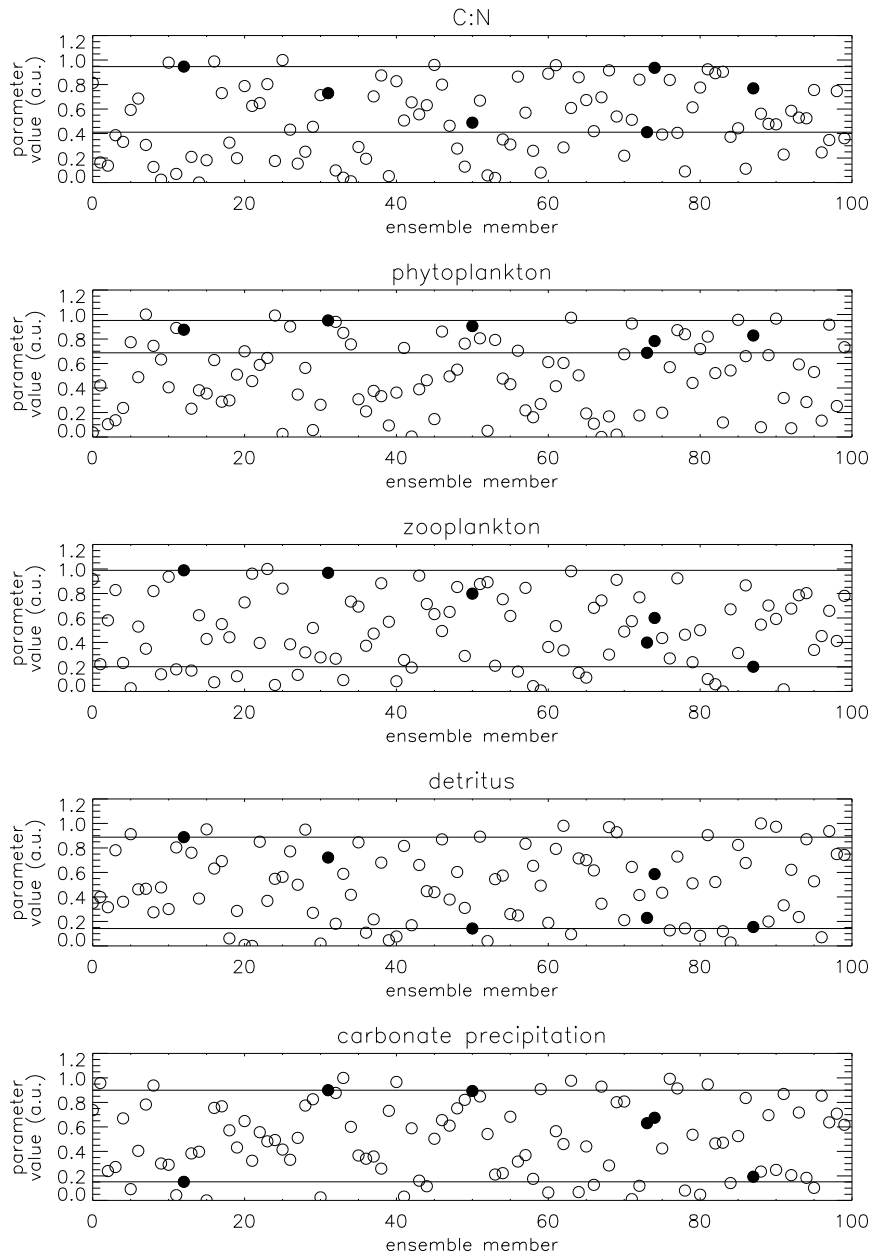


**Fig. 6.** Difference between the combined-PFT LAI of the mid-Holocene and  $\alpha 7$  runs (a–g) and the equivalent residual plots for the LGM and  $\alpha 7$  runs (h–n).

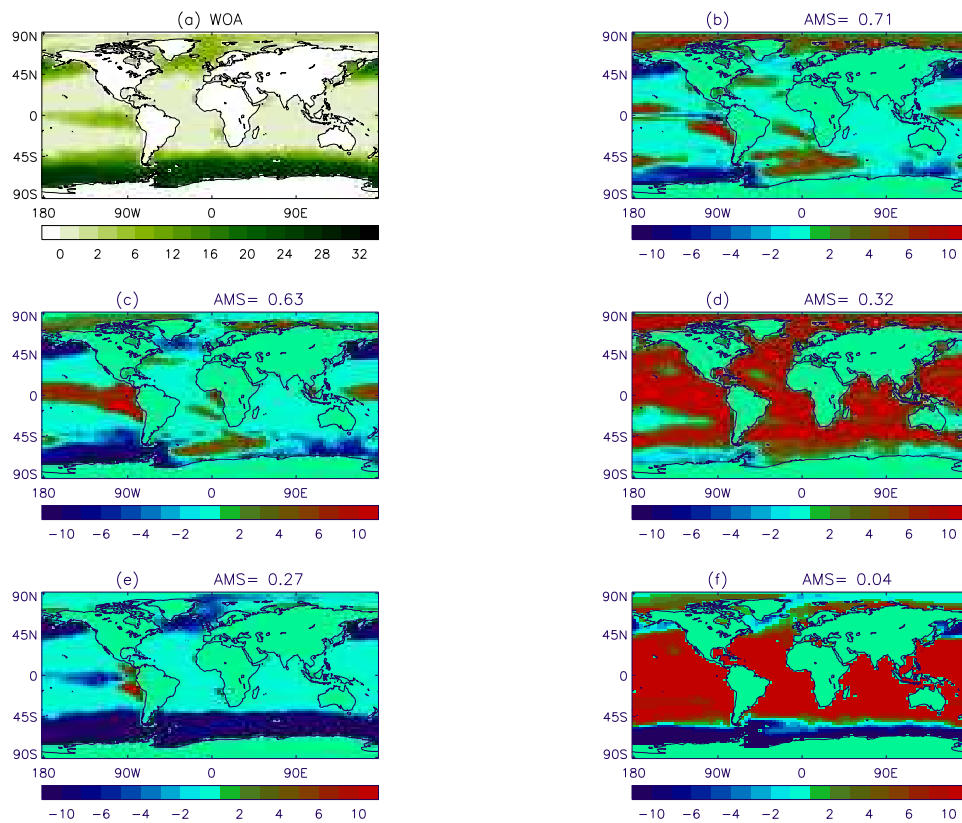


**Fig. 7.** The AMS for the ocean carbon cycle ensemble’s nitrate concentration when compared against World Ocean Atlas data. The ensemble member giving rise to the highest AMS is marked with a filled circle.

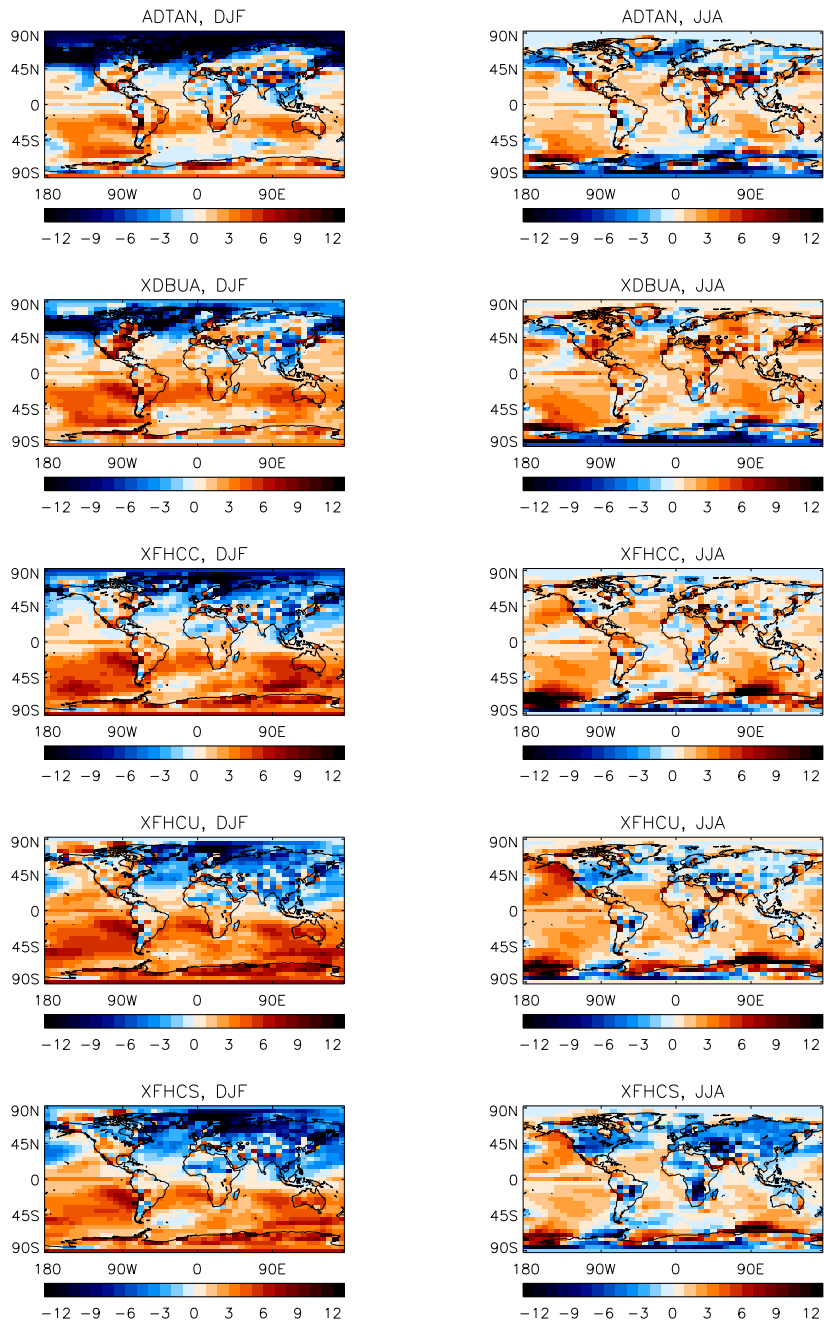




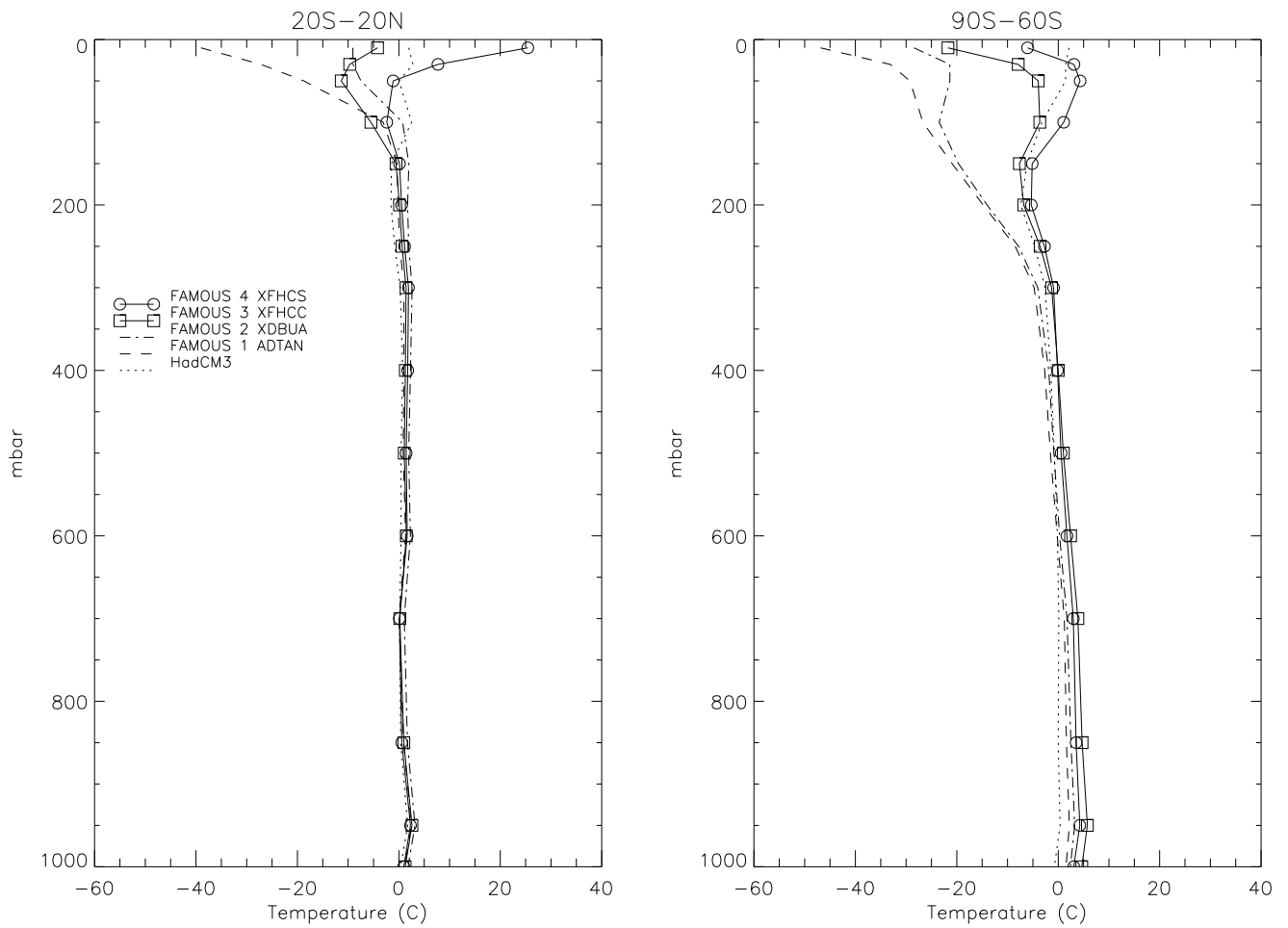
**Fig. 8.** The sensitivity of the ocean ensemble members to individual parameters. The top scoring six (in terms of their AMS) simulations are shown with filled symbols and the horizontal lines represent the minimum and maximum values of each parameter covered by them.



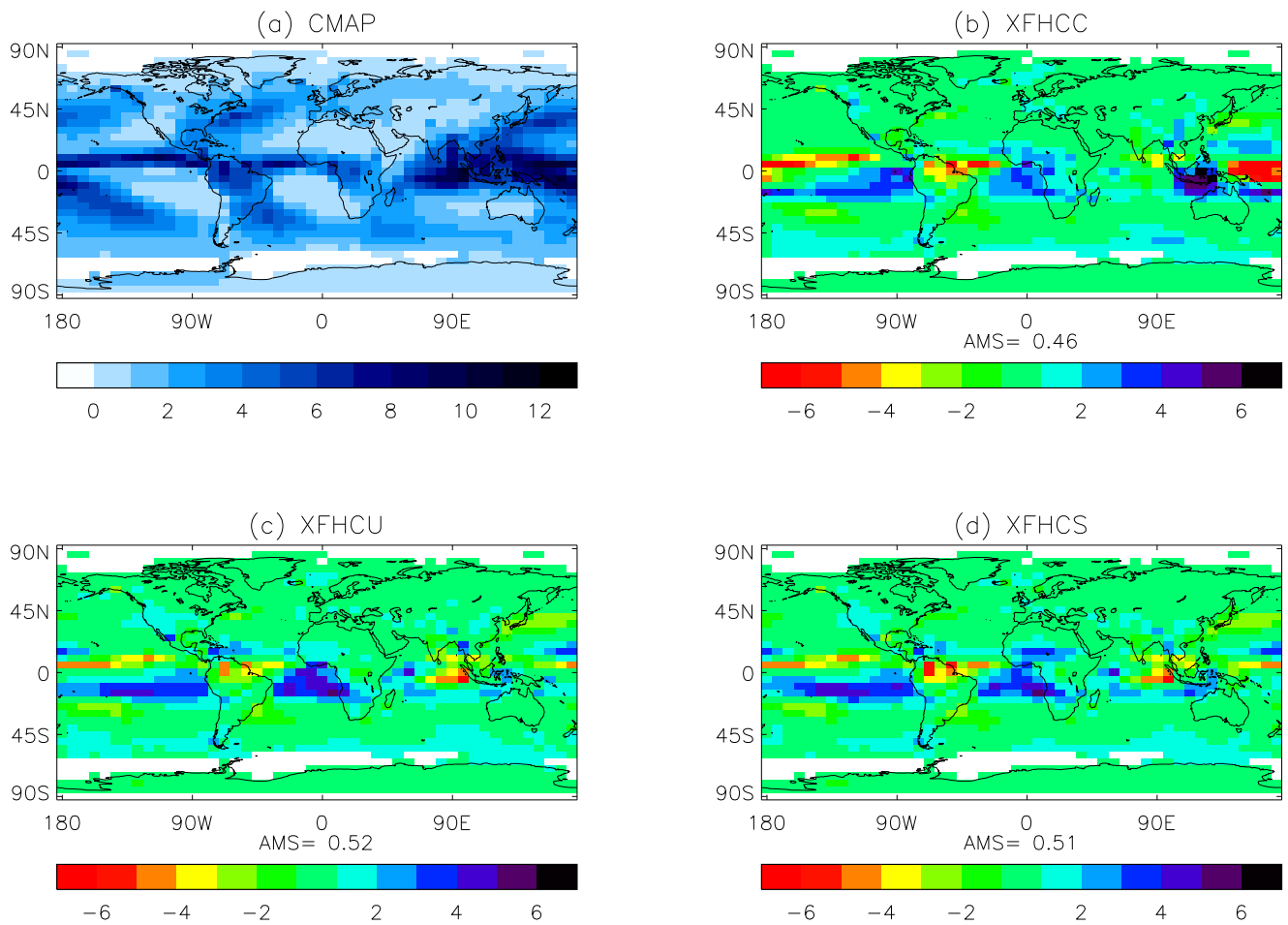
**Fig. 9.** Annual mean nitrate concentration in  $\text{mmol per m}^{-3}$  at 5 m for (a) World Ocean Atlas observations (Garcia et al., 2006) and (b–f) 5 ensemble members with AMS varying between the maximum and minimum for the perturbed physics ocean ensemble.



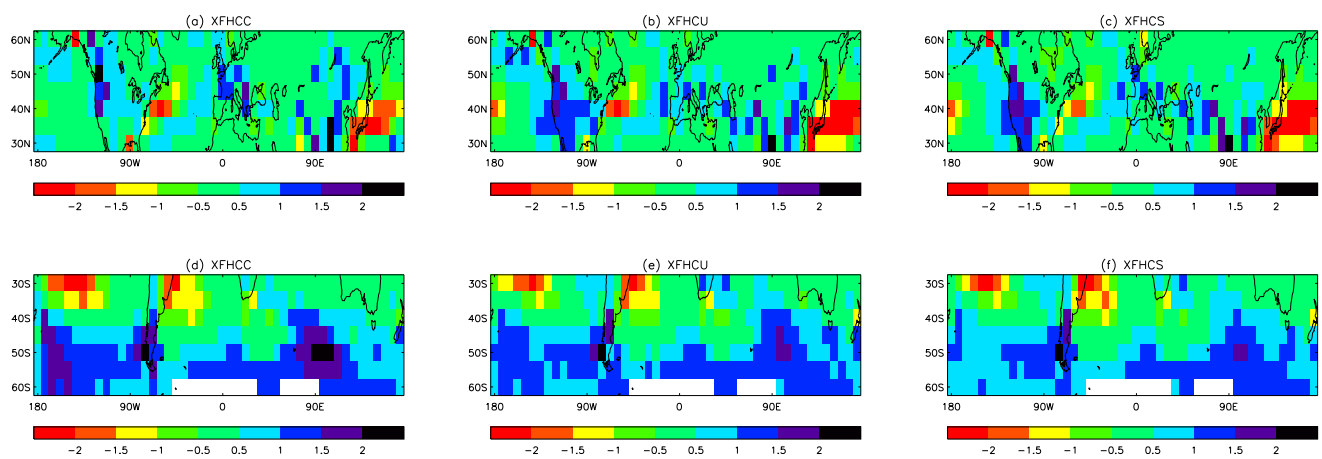
**Fig. 10.** Air temperature at 1.5 m with respect to HadCM3 for progressively more modern versions of FAMOUS (most recent at the bottom of the figure) for DJF (left) and JJA (right).



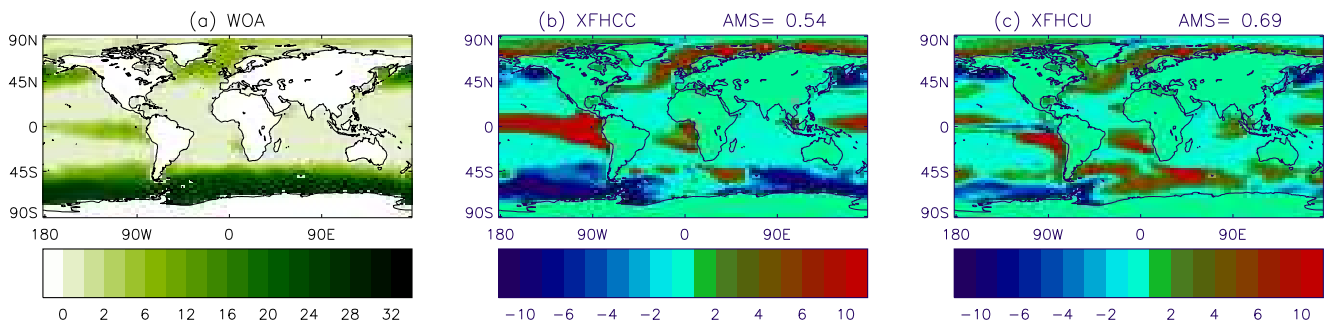
**Fig. 11.** Mean temperature profiles for 20° S–20° N (left) and 90° S–60° S (right) for HadCM3 (dotted line), generation 1 FAMOUS, ADTAN (dashed line), generation 2, XDBUA (dotted-dashed line), generation 3, XFHCC (squares) and generation 4b, XFHCS (circles). The generation 4a model, XFHCU is not shown since its temperature profile is virtually indistinguishable from the generation 4b version, especially in the stratosphere.



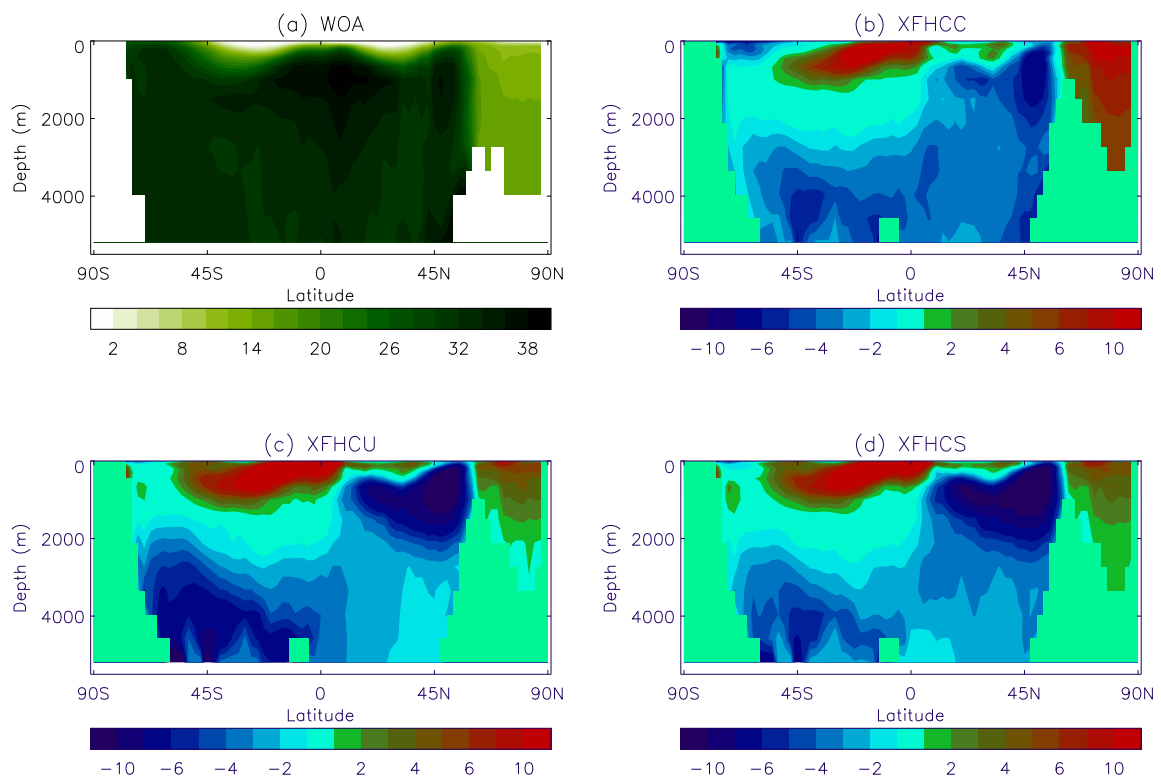
**Fig. 12.** Annual mean total precipitation rate in mm per day for (a) the CMAP climatology (Xie and Arkin, 1997) and the difference between the simulated total precipitation and CMAP for (b) the generation 3 model XFHCC, (c) the generation 4a model XFHCU and (d) the generation 4b model XFHCS. Missing data areas are set to white and the AMS scores for the 3 model generations are given in the subtitles to (b), (c) and (d).



**Fig. 13.** Difference between simulated and observed precipitation in the northern (a–c) and southern (d–f) mid-latitudes as shown globally in Fig. 12. Note the different contour intervals compared to Fig. 12.



**Fig. 14.** Annual mean nitrate concentration in mmol per m<sup>-3</sup> at 5 m for (a) World Ocean Atlas observations (Garcia et al., 2006) and the difference between the simulated and observed values for (b) the generation 3 model XFHCC, (c) the generation 4a model XFHCU. The AMS scores for simulations are given above (b) and (c).



**Fig. 15.** Zonal mean – depth plots for annual mean nitrate concentration in  $\text{mmol per m}^{-3}$  at 5 m for (a) World Ocean Atlas observations (Garcia et al., 2006) and the difference between the simulated and observed values for (b) the generation 3 model XFHCC, (c) the generation 4a model XFHCU, (d) the generation 4b model XFHCS.



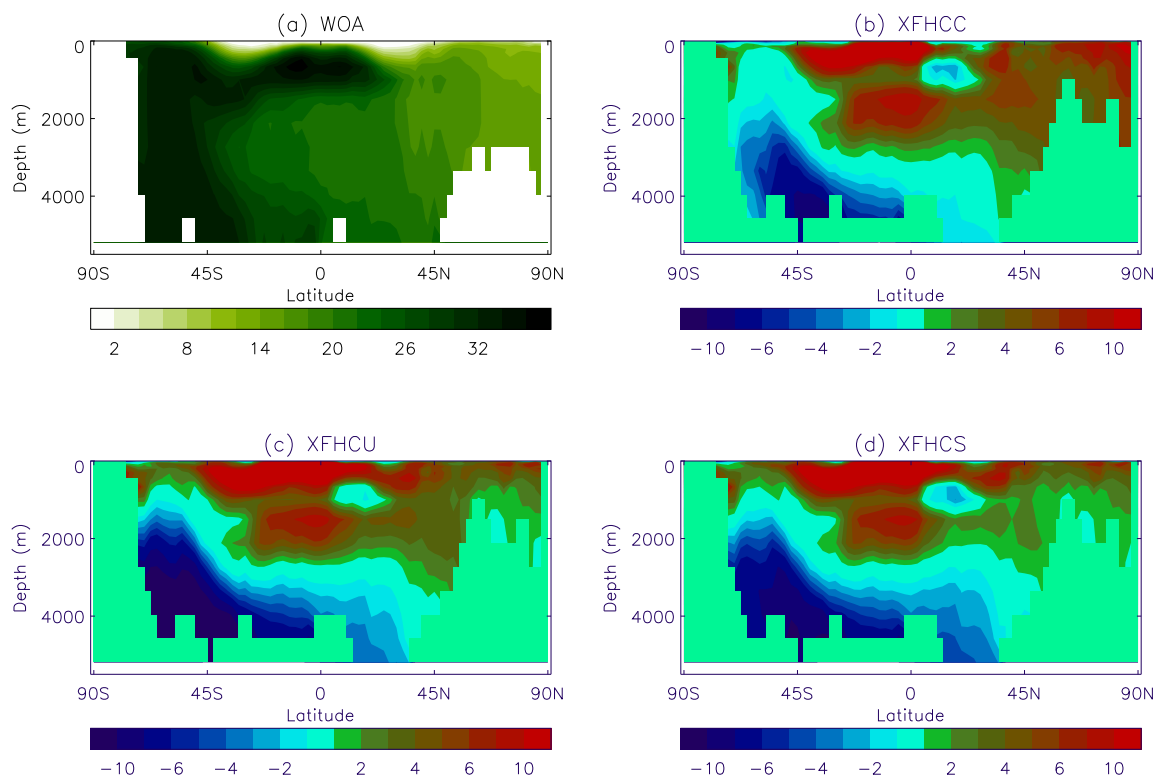


Fig. 16. As for Figure 15 but for the Atlantic basin ( $-70^{\circ}\text{W} - 20^{\circ}\text{E}$ ).

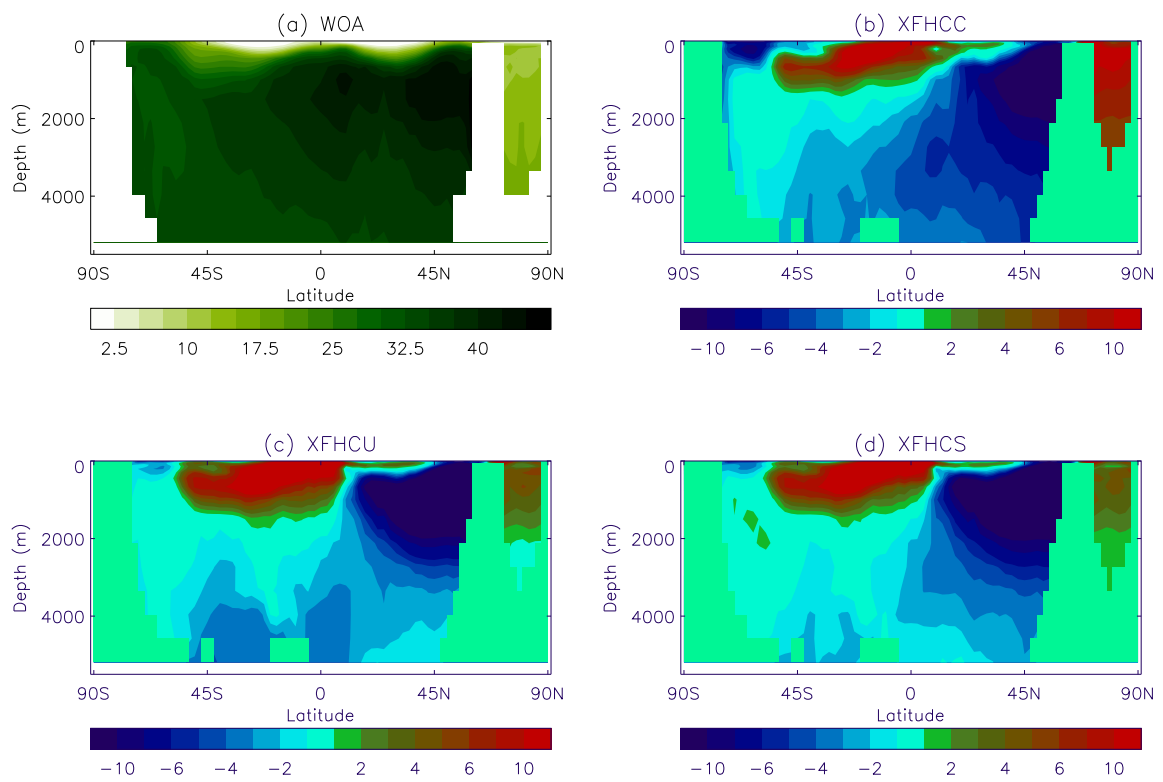
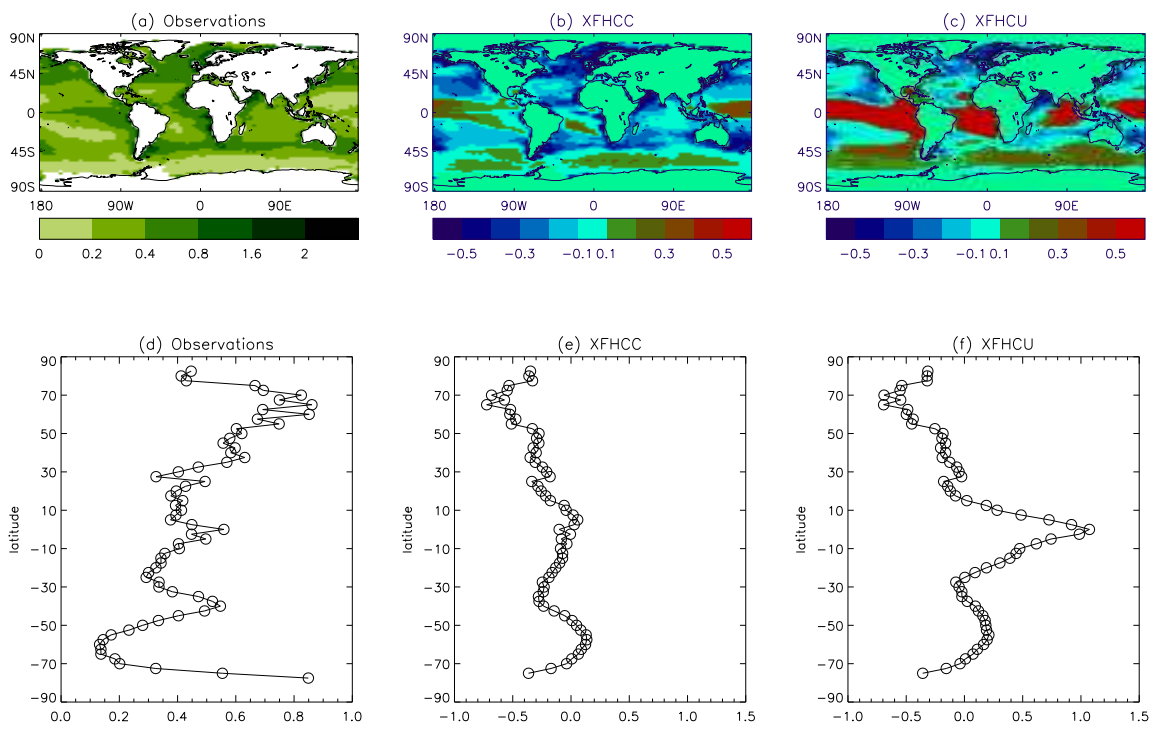


Fig. 17. As for Figure 15 but for the Pacific basin (150°E – 290°E).



**Fig. 18.** (a) Observed primary production (Behrenfeld and Falkowski, 1997) and (b–c) difference between the generation 3 model XFHC and the generation 4a model XFHCU and observational data. Sub-figures (d–f) show the same data but for zonally averaged quantities. The units of all quantities in this figure are gC per m<sup>2</sup> per day.